

Premier Reference Source

Quality Assurance and Assessment Practices in Translation and Interpreting

Copyright 2018, Information Science Reference. All rights reserved. May not be reproduced in any form without permission from the publisher, except flash users permitted under U.S. or applicable copyright law.



Elsa Huertas-Barros, Sonia Vandepitte,
and Emilia Iglesias-Fernández

IGI Global
DISSEMINATOR OF KNOWLEDGE

EBSCO Publishing (EBSCOhost) - printed on 2/10/2023 12:17 PM via
AN: 1842087 ; Huertas-Barros, Elsa, Vandepitte, Sonia, Iglesias-Fernandez, Emilia.; Quality Assurance and Assessment
Practices in Translation and Interpreting
Account: ns335141

Quality Assurance and Assessment Practices in Translation and Interpreting

Elsa Huertas-Barros
University of Westminster, UK

Sonia Vandepitte
Universiteit Gent, Belgium

Emilia Iglesias-Fernández
Universidad de Granada, Spain

A volume in the Advances in Linguistics and
Communication Studies (ALCS) Book Series



Published in the United States of America by

IGI Global
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA, USA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: <http://www.igi-global.com>

Copyright © 2019 by IGI Global. All rights reserved. No part of this publication may be reproduced, stored or distributed in any form or by any means, electronic or mechanical, including photocopying, without written permission from the publisher. Product or company names used in this set are for identification purposes only. Inclusion of the names of the products or companies does not indicate a claim of ownership by IGI Global of the trademark or registered trademark.

Library of Congress Cataloging-in-Publication Data

Names: Huertas-Barros, Elsa, 1984- editor. | Vandepitte, Sonia, editor. | Iglesias-Fernandez, Emilia, 1966- editor.

Title: Quality assurance and assessment practices in translation and interpreting / Elsa Huertas-Barros, Sonia Vandepitte, and Emilia Iglesias-Fernandez, editors.

Description: Hershey PA : Information Science Reference, [2018]

Identifiers: LCCN 2017037442 | ISBN 9781522552253 (hardcover) | ISBN 9781522552260 (ebook)

Subjects: LCSH: Translating and interpreting--Technological innovations. | Translating and interpreting--Study and teaching. | Translating and interpreting--Evaluation. | Translation studies.

Classification: LCC P306.97.T73 Q445 2018 | DDC 418/.020711--dc23 LC record available at <https://lcn.loc.gov/2017037442>

This book is published in the IGI Global book series Advances in Linguistics and Communication Studies (ALCS) (ISSN: 2372-109X; eISSN: 2372-1111)

British Cataloguing in Publication Data

A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

For electronic access to this publication, please contact: eresources@igi-global.com.



Advances in Linguistics and Communication Studies (ALCS) Book Series

Abigail G. Scheg
Western Governors University, USA

ISSN:2372-109X
EISSN:2372-1111

MISSION

The scope of language and communication is constantly changing as society evolves, new modes of communication are developed through technological advancement, and novel words enter our lexicon as the result of cultural change. Understanding how we communicate and use language is crucial in all industries and updated research is necessary in order to promote further knowledge in this field.

The **Advances in Linguistics and Communication Studies (ALCS)** book series presents the latest research in diverse topics relating to language and communication. Interdisciplinary in its coverage, ALCS presents comprehensive research on the use of language and communication in various industries including business, education, government, and healthcare.

COVERAGE

- Graphic Communications
- Dialectology
- Interpersonal Communication
- Forensic Linguistics
- Language and Identity
- Language in the Media
- Language Acquisition
- Youth Language
- Non-Verbal Communication
- Cross-Cultural Communication

IGI Global is currently accepting manuscripts for publication within this series. To submit a proposal for a volume in this series, please contact our Acquisition Editors at Acquisitions@igi-global.com or visit: <http://www.igi-global.com/publish/>.

The Advances in Linguistics and Communication Studies (ALCS) Book Series (ISSN 2372-109X) is published by IGI Global, 701 E. Chocolate Avenue, Hershey, PA 17033-1240, USA, www.igi-global.com. This series is composed of titles available for purchase individually; each title is edited to be contextually exclusive from any other title within the series. For pricing and ordering information please visit <http://www.igi-global.com/book-series/advances-linguistics-communication-studies/78950>. Postmaster: Send all address changes to above address. Copyright © 2019 IGI Global. All rights, including translation in other languages reserved by the publisher. No part of this series may be reproduced or used in any form or by any means – graphics, electronic, or mechanical, including photocopying, recording, taping, or information and retrieval systems – without written permission from the publisher, except for non commercial, educational use, including classroom teaching purposes. The views expressed in this series are those of the authors, but not necessarily of IGI Global.

Titles in this Series

For a list of additional titles in this series, please visit: www.igi-global.com/book-series

Establishing and Evaluating Digital Ethos and Online Credibility

Moe Folk (Kutztown University of Pennsylvania, USA) and Shawn Apostel (Bellarmine University, USA)
Information Science Reference • copyright 2017 • 400pp • H/C (ISBN: 9781522510727) • US \$200.00 (our price)

Computational and Cognitive Approaches to Narratology

Takashi Ogata (Iwate Prefectural University, Japan) and Taisuke Akimoto (The University of Electro-Communications, Japan)
Information Science Reference • copyright 2016 • 467pp • H/C (ISBN: 9781522504320) • US \$215.00 (our price)

Analyzing Language and Humor in Online Communication

Rotimi Taiwo (Obafemi Awolowo University, Nigeria) Akinola Odeunmi (University of Ibadan, Nigeria) and Akin Adetunji (Emmanuel Alayande College of Education, Nigeria)
Information Science Reference • copyright 2016 • 324pp • H/C (ISBN: 9781522503385) • US \$185.00 (our price)

Discourse Analysis as a Tool for Understanding Gender Identity, Representation, and Equality

Nazmunnessa Mahtab (University of Dhaka, Bangladesh) Sara Parker (John Moores University, UK) Farah Kabir (Action Aid, Bangladesh) Tania Haque (University of Dhaka, Bangladesh) Aditi Sabur (University of Dhaka, Bangladesh) and Abu Saleh Mohammad Sowad (University of Dhaka, Bangladesh)
Information Science Reference • copyright 2016 • 323pp • H/C (ISBN: 9781522502258) • US \$185.00 (our price)

Revealing Gender Inequalities and Perceptions in South Asian Countries through Discourse Analysis

Nazmunnessa Mahtab (University of Dhaka, Bangladesh) Sara Parker (John Moores University, UK) Farah Kabir (Action Aid, Bangladesh) Tania Haque (University of Dhaka, Bangladesh) Aditi Sabur (University of Dhaka, Bangladesh) and Abu Saleh Mohammad Sowad (University of Dhaka, Bangladesh)
Information Science Reference • copyright 2016 • 286pp • H/C (ISBN: 9781522502791) • US \$185.00 (our price)

Interdisciplinary Perspectives on Contemporary Conflict Resolution

Paulo Novais (University of Minho, Portugal) and Davide Carneiro (University of Minho, Portugal)
Information Science Reference • copyright 2016 • 363pp • H/C (ISBN: 9781522502456) • US \$185.00 (our price)

Handbook of Research on Effective Communication, Leadership, and Conflict Resolution

Anthony H. Normore (California State University Dominguez Hills, USA & International Academy of Public Safety, USA) Larry W. Long (Illinois State University, USA & International Academy of Public Safety, USA) and Mitch Javidi (North Carolina State University, USA & International Academy of Public Safety, USA)
Information Science Reference • copyright 2016 • 733pp • H/C (ISBN: 9781466699700) • US \$315.00 (our price)



701 East Chocolate Avenue, Hershey, PA 17033, USA
Tel: 717-533-8845 x100 • Fax: 717-533-8661
E-Mail: cust@igi-global.com • www.igi-global.com

Editorial Advisory Board

Maria González Davies, *Universitat Ramon Llull, Spain*

Dorothy Kenny, *Dublin City University, Ireland*

Geoffrey Koby, *Kent State University, USA*

Sara Ramos Pinto, *University of Leeds, UK*

Mariachiara Russo, *Università di Bologna, Italy*

Marina Sánchez Torrón, *The University of Auckland, New Zealand*

Elisabet Tiselius, *Stockholm University, Sweden*

Roelof van Deemter, *Rijksuniversiteit Groningen, The Netherlands*

Catherine Way, *Universidad de Granada, Spain*

List of Reviewers

Joke Daems, *Ghent University, Belgium*

Joanna Gough, *University of Surrey, UK*

Kyriaki Kourouni, *Aristotle University, Greece*

Xiangdong Li, *Xi'an International Studies University, China*

Lieve Macken, *Ghent University, Belgium*

Maria Piotrowska, *Jagiellonian University, Poland*

Mirjam Sepesy Maučec, *University of Maribor, Slovenia*

Izabel Emilia Telles de Vasconcelos Souza, *Osaka University, Japan*

Caiwen Wang, *University of Westminster, UK*

Table of Contents

Foreword	XV
Preface	xvii
Acknowledgment	XXX

Section 1

Introduction to the Field of Translation Quality Assessment

Chapter 1

Quality and Quality Assessment in Translation: Paradigms in Perspective	1
<i>Marcel Thelen, Maastricht School of Translation and Interpreting, The Netherlands</i>	

Chapter 2

How to Put the Translation Test to the Test? On Preselected Items Evaluation and Perturbation	26
<i>Gys-Walt Van Egdom, Zuyd University of Applied Sciences, The Netherlands</i>	
<i>Heidi Verplaetse, KU Leuven, Belgium</i>	
<i>Iris Schrijver, University of Antwerp, Belgium</i>	
<i>Hendrik J. Kockaert, KU Leuven, Belgium</i>	
<i>Winibert Segers, KU Leuven, Belgium</i>	
<i>Jasper Pauwels, University of Antwerp, The Netherlands</i>	
<i>Bert Wylin, KU Leuven, Belgium</i>	
<i>Henri Bloemen, KU Leuven, Belgium</i>	

Chapter 3

Quality Assurance in Translation and Corporate Communications: Exploring an Interdisciplinary Interface	57
<i>Gary Massey, Zurich University of Applied Sciences, Switzerland</i>	
<i>Regine Wieder, Zurich University of Applied Sciences, Switzerland</i>	

Section 2

Translation Quality Assessment in Interpreting and Audio-Visual Translation

Chapter 4

Conceptualizing and Operationalizing a Formative Assessment Model for English-Chinese Consecutive Interpreting: A Case Study in an Undergraduate Interpreting Course	89
<i>Chao Han, Southwest University, China</i>	

Chapter 5	
Competency-Based Education and Assessment: A Proposal for US Court Interpreter Certification	112
<i>Melissa Wallace, University of Texas at San Antonio, USA</i>	

Chapter 6	
Dischronies and the Absence of Image in the Evaluation of Dubbing	133
<i>José Tomás Conde Ruano, University of the Basque Country, Spain</i>	

Chapter 7	
Quality Assessment in Audio Description: Lessons Learned From Interpreting	155
<i>Louise Fryer, University College London, UK</i>	

Section 3
Process-Oriented Translation Quality Assessment

Chapter 8	
Process-Oriented Assessment of Problems and Errors in Translation: Expanding Horizons Through Screen Recording	179
<i>Erik Angelone, Kent State University, USA</i>	

Chapter 9	
Exploring Process-Oriented Translation Competence Assessment: Rationale, Necessity, and Potential From a Problem-Solving Perspective	199
<i>Si Cheng, The University of Manchester, UK</i>	

Chapter 10	
Cognitive Effort and Efficiency in Translation Revision	226
<i>Moritz Schaeffer, Johannes Gutenberg-Universität Mainz, Germany</i>	
<i>Anke Tardel, Johannes Gutenberg-Universität Mainz, Germany</i>	
<i>Sascha Hofmann, Johannes Gutenberg-Universität Mainz, Germany</i>	
<i>Silvia Hansen-Schirra, Johannes Gutenberg-Universität Mainz, Germany</i>	

Section 4
Learning Translation Quality Assessment

Chapter 11	
Constructing Standards in Communities: Tutors' and Students' Perceptions of Assessment Practices on an MA Translation Course	245
<i>Elsa Huertas-Barros, University of Westminster, UK</i>	
<i>Juliet Vine, University of Westminster, UK</i>	

Chapter 12	
Introducing a Student Self-Evaluation Grid for Translation Assignments	270
<i>Mari J. Pakkala-Weckström, University of Helsinki, Finland</i>	

Chapter 13	
Assessing Translation Students' Reflective and Autonomous Learning.....	288
<i>Ya-Yun Chen, Newcastle University, UK</i>	
Chapter 14	
The Role of Expertise in Peer Feedback Analysis: Exploring Variables and Factors in a Translation Context.....	315
<i>Sonia Vandepitte, Ghent University, Belgium</i>	
<i>Joleen Hanson, University of Wisconsin – Stout, USA</i>	
Chapter 15	
Student Peer Feedback in a Translation Task: Experiences With Questionnaires and Focus Group Interviews.....	337
<i>Carmen Heine, Aarhus University, Denmark</i>	
Compilation of References	358
About the Contributors	401
Index	406

Detailed Table of Contents

Foreword	XV
Preface	xvii
Acknowledgment	XXX

Section 1

Introduction to the Field of Translation Quality Assessment

Chapter 1

Quality and Quality Assessment in Translation: Paradigms in Perspective	1
<i>Marcel Thelen, Maastricht School of Translation and Interpreting, The Netherlands</i>	

This chapter first gives an overview of existing interpretations of the concept of quality in organizations in general, as well as in the translation industry and translator training. This is followed by a discussion of quality assessment and how this is generally dealt with in the translation industry and translator training. Various perspectives are possible, but the focus is on translation as service provision as outlined in the translation services standards EN 15038 and ISO 17100. It is argued that implementing quality assessment of translation as service provision in translator training requires the introduction into the curriculum of quality management and quality assurance by students; a so-called skills lab is an ideal environment for this. Various suggestions are then given for assessing quality management and quality assurance while taking into account the requirements of validity and reliability.

Chapter 2

How to Put the Translation Test to the Test? On Preselected Items Evaluation and Perturbation	26
<i>Gys-Walt Van Egdome, Zuyd University of Applied Sciences, The Netherlands</i>	
<i>Heidi Verplaetse, KU Leuven, Belgium</i>	
<i>Iris Schrijver, University of Antwerp, Belgium</i>	
<i>Hendrik J. Kockaert, KU Leuven, Belgium</i>	
<i>Winibert Segers, KU Leuven, Belgium</i>	
<i>Jasper Pauwels, University of Antwerp, The Netherlands</i>	
<i>Bert Wylin, KU Leuven, Belgium</i>	
<i>Henri Bloemen, KU Leuven, Belgium</i>	

Reliable and valid evaluation of translation quality is one of the fundamental thrusts in present-day applied translation studies. In this chapter, a thumbnail sketch is provided of the developments, in and outside of translation studies, that have contributed to the ubiquity of quality in translation discourse.

This sketch reveals that we will probably never stand poised to reliably and validly measure the quality of translation in all its complexity and its ramifications. Therefore, the authors have only sought to address the issue of product quality evaluation. After an introduction of evaluation methods, the authors present the preselected items evaluation method (PIE method) as a perturbative testing technique developed to evaluate the quality of the target text (TT). This presentation is flanked by a case study that has been carried out at the University of Antwerp, KU Leuven, and Zuyd University of Applied Sciences. The case study shows that, on account of its perturbative qualities, PIE allows for more reliable and more valid measurement of product quality.

Chapter 3

Quality Assurance in Translation and Corporate Communications: Exploring an Interdisciplinary Interface 57

Gary Massey, Zurich University of Applied Sciences, Switzerland

Regine Wieder, Zurich University of Applied Sciences, Switzerland

While the nature and status of translators' work are changing due to technologisation and other factors, translation is acquiring a strategic function in organisations. The intercultural component of translation competence makes translators well positioned to play a key role in assuring quality in international corporate communications. But quality models envisage only restricted interactions between translators, clients and communications specialists. Moreover, evidence about translators' self-concepts shows them underequipped to adopt the roles that meaningful cooperation with corporate communications suggests. This chapter reports on a pilot study at the interface between translation and corporate communications in Switzerland. Presenting findings from a survey of translation and communications professionals, it reveals underdeveloped feedforward and feedback cultures and a translator self-concept that underplays the mediatory, advisory added value of human translation. Concrete implications for quality assurance and translator education are drawn and future research is outlined.

Section 2

Translation Quality Assessment in Interpreting and Audio-Visual Translation

Chapter 4

Conceptualizing and Operationalizing a Formative Assessment Model for English-Chinese Consecutive Interpreting: A Case Study in an Undergraduate Interpreting Course 89

Chao Han, Southwest University, China

Formative assessment has been increasingly used by interpreter trainers and educators to promote student learning. Different forms of formative assessment have been practiced and reported in interpreting literature. However, a critical review of current practices reported in literature suggests that a longitudinally designed formative assessment model that harnesses the synergistic potential of self, peer, and teacher assessment seems to be lacking. This chapter therefore aims to provide a detailed account of how an inclusive formative assessment model was conceptualized and operationalized for an undergraduate-level English-Chinese consecutive interpreting course and how students and the teacher perceived the assessment model. Based on the students' evaluation and the teacher's reflection, the chapter highlights good practices that contribute to effective formative assessment, discusses potential problems, proposes possible solutions, and suggests future trends in implementing and researching formative assessment in interpreter training and education.

Chapter 5

Competency-Based Education and Assessment: A Proposal for US Court Interpreter

Certification 112

Melissa Wallace, University of Texas at San Antonio, USA

In an attempt to analyze the reliability and validity of the most frequently used oral certification exams for court interpreters in the United States, this chapter examines the basic test model used for state-level certification through the lens of concepts in testing theory. Having identified several limitations to the currently used performance-based model, a hybrid model which includes competency-based education and assessment is proposed. By building on best practices in competency-based education, the alternative credentialing paradigm proposed here would represent an innovation in the context of court interpreter certification in the United States, requiring the transfer of assessment criteria usually used in traditional educational contexts into the realm of professional training. The proposed hybrid model would necessitate a shift from one high-stakes exam to assessment of a series of compartmentalized competency clusters that would account for soft skills and dispositional traits not currently assessed.

Chapter 6

Dischronies and the Absence of Image in the Evaluation of Dubbing 133

José Tomás Conde Ruano, University of the Basque Country, Spain

An approximation of quality in audiovisual translation is presented, based upon an experiment carried out with students of different level of expertise, in order to check whether considering the image affects the evaluation of quality in dubbing and which dischrony has a greater impact on students' ratings. When they watch the scenes on video, evaluators (i.e. the students) bestow them higher ratings as opposed to when they have only the text. The lack of synchronization, which is penalized the most, is lip synchronization, although the differences do not seem to be significant. The average ratings given by the two groups of students are similar, but there is a stronger coincidence among advanced students. The presentation on video beclouds potential linguistic errors made during the translation phase. Errors related to synchronization are not as relevant to the receivers' judgement as expected.

Chapter 7

Quality Assessment in Audio Description: Lessons Learned From Interpreting 155

Louise Fryer, University College London, UK

Audio description (AD) is one of the younger modes of translation. It shares many similarities with interpreting, although AD users have specific needs because they are blind or partially sighted. As quality is of concern in both fields, this chapter explores the overlaps to see what can be learned for AD from research already carried out in interpreting. Macro and micro criteria suggested for each discipline are compared, and describer competencies are discussed in the context of AdlabPRO, a European research project that seeks to define the professional profile of an audio describer and develop training materials and courses. The chapter concludes that assessment protocols and rating scales developed for interpreting might be adopted for AD, after appropriate adaptation to accommodate areas where the fit is incomplete. These include synchrony and the need for the AD to be considered, not in isolation, but in relation to the existing audio elements of the source text (ST).

Section 3 Process-Oriented Translation Quality Assessment

Chapter 8

Process-Oriented Assessment of Problems and Errors in Translation: Expanding Horizons Through Screen Recording	179
<i>Erik Angelone, Kent State University, USA</i>	

Screen recording has gradually emerged as an efficacious tool in the context of process-oriented translator training. From an assessment standpoint, process protocols derived from screen recordings would seem to hold great potential as an empirical means through which translators and translation trainers can re-trace errors found in translation products back to underlying problem triggers that emerge during their creation. This chapter will begin by outlining how screen recordings can be utilized to reverse engineer translation products for purposes of process-oriented assessment. A series of directly observable indicators will be linked with various error classification parameters, including locus (comprehension, transfer, or production), phase (drafting or revision), and information retrieval type (internal or external) in providing assessors with a diagnostic gauge for pinpointing potential triggers. The chapter will conclude with some preliminary data on evidence of inter-rater consistency when screen recording is used in such a diagnostic capacity by various student populations.

Chapter 9

Exploring Process-Oriented Translation Competence Assessment: Rationale, Necessity, and Potential From a Problem-Solving Perspective	199
<i>Si Cheng, The University of Manchester, UK</i>	

In translation education and training, competence assessment plays an essential role in fostering students' competence development. This chapter promotes process-oriented assessment from a translation problem-solving perspective, with supporting evidence from a longitudinal study. Although process-oriented competence assessment is increasingly becoming more important in translation education and training, the relation between the development of translation competence and the translation process remains underexplored. This chapter provides a translation problem-solving perspective to understand the rationale for and the necessity of process-oriented competence assessment and suggests practical and feasible process-oriented assessment tools in the translation classroom.

Chapter 10

Cognitive Effort and Efficiency in Translation Revision	226
<i>Moritz Schaeffer, Johannes Gutenberg-Universität Mainz, Germany</i>	
<i>Anke Tardel, Johannes Gutenberg-Universität Mainz, Germany</i>	
<i>Sascha Hofmann, Johannes Gutenberg-Universität Mainz, Germany</i>	
<i>Silvia Hansen-Schirra, Johannes Gutenberg-Universität Mainz, Germany</i>	

Empirical studies of revision are often based on either think aloud protocols, interviews, or observational methods. Eye tracking and keylogging methods are rarely applied to the study of revision behavior. The authors employ established methods from translation process research (TPR) to study the eye movement and typing behavior during self-revision (i.e., the phase in the translation process that follows a first complete draft). The authors measure the effect of behavior during the drafting phase on the relative revision duration. Relative revision duration is the time translators spend revising the first complete draft

of the source text. They find that the most efficient process involves a large degree of concurrent reading and writing and few deletions during the drafting phase. The efficiency gains in terms of relative revision duration achieved by avoiding discontinuous typing, by making a larger number of deletions, pausing for longer amounts of time, and engaging in less concurrent reading and writing are outweighed by the gains in total task time by doing the exact opposite.

Section 4 **Learning Translation Quality Assessment**

Chapter 11

Constructing Standards in Communities: Tutors' and Students' Perceptions of Assessment Practices on an MA Translation Course	245
<i>Elsa Huertas-Barros, University of Westminster, UK</i>	
<i>Juliet Vine, University of Westminster, UK</i>	

Assessment practices on translation programs provide a valuable lens through which to view current understandings about the nature of translation pedagogy. In the context of competence-based training, the last decade has seen the proliferation of assessment instruments aimed at enhancing students' learning by prioritising competence development and the translation process. Using the University of Westminster as a case study, the authors have sought to provide a clearer insight into the current understandings of translation and assessment practices on the MA Translation courses in the light of the current debates in translation pedagogy. The authors undertook a two-pronged approach by surveying not only the tutors, but also the students. This chapter will present and analyse the findings of the two surveys on assessment practices using the framework of the six tenets of good assessment practice set out by the Higher Education Academy (HEA) and, in particular, assess to what extent assessment literacy has been developed.

Chapter 12

Introducing a Student Self-Evaluation Grid for Translation Assignments	270
<i>Mari J. Pakkala-Weckström, University of Helsinki, Finland</i>	

This chapter will report the results of a study introducing a student self-evaluation grid for translation assignments, based on previous work by Marc Orlando. The grid described here was developed with and for second-year students of English Translation at the University of Helsinki during the autumn terms of 2015 and 2016. This process and the results are described in the light of a pilot study conducted with the students. Based on student feedback, the grid seems to provide a structured framework for evaluating both one's translation process and the translation product, but there are also areas to be developed in this system.

Chapter 13

Assessing Translation Students' Reflective and Autonomous Learning.....	288
<i>Ya-Yun Chen, Newcastle University, UK</i>	

As an essential soft skill in life-long learning, reflective and autonomous learning has been an integral part of many translator training programs today. However, how it could be assessed systematically and what factors might influence its acquisition is still much under-researched. To help bridge this gap, this chapter aims at reporting the findings of an empirical study, which used diary, think-aloud and small-group discussion as reflective learning methods to investigate translation students' reflective learning. It first

provides an overview of relevant theory and then reports how students' reflective levels were assessed and analyzed quantitatively and qualitatively. Based on the empirical findings, it discusses the factors influencing the success of a reflection-encouraging learning environment, followed by a provisional model of translation students' reflective and autonomous learning process.

Chapter 14

The Role of Expertise in Peer Feedback Analysis: Exploring Variables and Factors in a Translation Context..... 315

Sonia Vandepitte, Ghent University, Belgium
Joleen Hanson, University of Wisconsin – Stout, USA

Since Kiraly pointed out the beneficial role of collaboration in translation training, increasingly more attention has been paid to the potential benefits of peer collaboration. While Wang and Han studied translation trainees' explicit perceptions of any benefits resulting from peer feedback, the present contribution first investigates the role of translator's implicit perceptions of reviewer expertise in the effectiveness of a peer comment in a case study. It then inquires into the number and type of peer feedback comments in relation to whether the target language that is to be reviewed is the reviewer's L1 or L2. Here, two data sets are hypothesized to yield similar results: (1) a set of native and non-native reviewer comments and (2) the comments written by translators in a direct translation situation and in an inverse translation situation. Findings were surprising, however, and professional, methodological, and theoretical research implications for translation and revision competence models are suggested.

Chapter 15

Student Peer Feedback in a Translation Task: Experiences With Questionnaires and Focus Group Interviews..... 337

Carmen Heine, Aarhus University, Denmark

Social science research methods can help shed light on students' peer feedback performance. They can also help enlighten researchers on students' reception and repercussion to feedback tasks. The operationalizability of these methods for future peer activities in Translation Didactics is examined in this chapter. Multiple peer feedback data from undergraduate Business Communication students is compared with questionnaire and interview data. The data derives from peer feedback loops and provides insights into the students' perception of working with peer feedback on a web-text localization and translation commentary task performed to mirror professional practice. The analysis of the wording of student feedback, of the revisions suggested and the analysis of whether or not—and how—students implement their peer's suggestions, allows qualitative evaluation and interpretation. The methods applied are compared and their feasibility for further research into peer feedback in Translation Studies is explored.

Compilation of References 358

About the Contributors 401

Index..... 406

Foreword

Since the first translation of the Bible to interpreting at the Nuremberg trial, cultural mediators have been facilitating communication between different groups of people in different ways across the globe for centuries. Linking different cultures and languages has later become a profession in its own right which, later on, had to adapt to the digitalisation of the world.

Technology has changed the way we think, we live, we communicate, and we work. For translators and interpreters, technology has shaped their profession in such a radical way that the skills they need in order to perform keep changing as well. This varies from access to information via the internet to machine translation, translation memories, speech processing or dubbing software.

Training translators and interpreters has become a huge challenge for academia which has to keep up with technological evolution in order to make sure that the training it provides fulfils the industry's requests. It faces the challenge to train translators and interpreters who can efficiently adapt to the needs of a constantly changing industry.

The European Commission is one of the largest employers of translators and interpreters in the world. It provides translation of EU legislation in 24 official languages and makes sure that EU stakeholders who meet daily are able to communicate between each other in spite of their language differences. It is therefore in its interest to follow closely developments in translation and interpreter training in Europe and to contribute to them, in order to make sure that there is an adequate supply of highly qualified translators and interpreters available to meet its own requirements but also those of a wider translation and interpreting market.

In this context, quality assessment has become increasingly important whether it is for training purposes, for evaluating a final product or for analysing the very process of translating or interpreting. It encompasses a wide range of translation and interpreting practices – from institutional translation, interpreting, project management, audiovisual translation, audiodescription or dubbing.

This volume addresses the complex issue of quality assessment both in a formative environment (academia) and a professional one (industry) and provides valuable links between the two. The translation process is broken down into several stages or areas of competences, allowing for the translation competence to be used as a framework for assessing translation. This enables the constant evaluation of trainees' skills making sure they acquire adequate assessment literacy.

The editors tackle the challenge of bringing together both worlds by providing a platform for reflection and further enhancement. This volume is an essential contribution to the translator and interpreter profession from which both the academia and the industry will undoubtedly benefit.

Angeliki Petrits
European Commission

Angeliki Petrits is currently a policy coordinator with the European Commission. As a staff translator for over 25 years, she has managed several projects related to Machine Translation and the European Master's in Translation (EMT) network which she helped set up in 2009. Her positions included a secondment to the London office where she was involved in a range of initiatives to promote language learning and translation in the UK and in raising awareness of opportunities for linguists within the European institutions. In 2016 she was awarded the Threlford Cup from the Chartered Institute of Linguists as an acknowledgement of her activities in the EMT network and her support to the UK universities. Angeliki has a BA in French from the University of Thessaloniki (1984), an MA (1985) and a PhD (1989) in Sociolinguistics from the University of Paris 7 and has published in the fields of sociolinguistics and translation.

Preface

Since translation and interpreting established themselves as professions and as academic disciplines, both the industry and the academic settings have evolved swiftly as a consequence of the significant changes affecting the field (Drugan, 2013, pp. 185; Saldanha & O'Brien, 2014, pp. 95) and the innovative approaches and concepts linked to the disciplines in recent decades (e.g., Romero-Trillo, 2015). In the workplace, the development of translation memories and machine translation have led to new translation quality assurance practices where translators have found themselves checking not only human translation, but also machine translation outputs. And in training settings, the new developments have inevitably resulted in new forms of feedback and assessment that are replacing more traditional ways to judge students' performance in translation and interpreting training (Huertas Barros & Vine, 2018). They include, for instance: diagnostic, summative and formative assessment, self-assessment, reflective diaries, translation commentaries and formative feedback by means of peer and self-assessment tasks. In this context, the notions of revision and interpersonal competences have gained great importance, with international projects such as OPTIMALE recognizing them as high priorities in the labor market, and many translation scholars calling upon revision training and the introduction of collaborative learning in translation education and training (e.g., Hurtado Albir, 1999/2003, 2007, 2015; Kiraly, 2000; González Davies, 2004; Kelly, 2005; Klimkowski, 2006; Way, 2008, 2014, 2016; Huertas Barros, 2011, 2013; Galán Mañas & Hurtado Albir, 2015; Lisaité et al., 2016).

Recent translation studies are exploring the notion of the peer feedback as a form of collaboration and its positive impact on translation competences (Lisaité et al., 2016; Vandepitte & Lisaité, 2016; Flanagan & Heine, 2015, 2017). From this perspective, providing effective peer feedback means incorporating Translation Quality Assessment into teaching, where practices of revision can be linked to a feedback process in the industry (i.e. students are introduced to professional quality standards, quality control criteria and benchmarks recognized at international level). The ongoing research project "Establishing competence levels in translation competence acquisition (written translation)" carried out by PACTE can also be seen as a first but solid step in this direction, as it will serve as a guide towards the establishment of criteria for professional quality control. Quality assessment plays, therefore, an essential role in both professional and academic settings. In the industry context, it is mainly linked to the quality of the translation and interpreting products and services. In education and training, quality assessment has two main roles, i.e. focusing on the translation and interpreting processes and on trainees' learning needs (formative function) and evaluating the knowledge acquired or grading students' achievements (summative function).

Quality is also a central notion in interpreter education, and Interpreting Quality Assessment is one of the most robust and prosperous fields in Interpreting Studies. From its outset, Interpreting Quality As-

assessment has been concerned with identifying a set of verbal and nonverbal criteria (e.g., Bühler, 1986; Kurz, 1993/2002) and determining their weight in the evaluation of both conference interpretation and interpreters. The importance that different groups of interpreting users attach to certain criteria (Gile, 1991; Kurz & Pöchhacker, 1995; Chiaro & Nocella, 2004; Collados Aís, Pradas Macías, Stévaux, & García Becerra, 2007; Zwischenberger & Pöchhacker, 2010; Collados Aís, Iglesias Fernández, Pradas Macías, & Stévaux, 2011) is useful in informing the design and development of consistent criteria. But findings show that rating criteria are difficult to separate (Collados Aís, 1998/2002; Pradas Macías, 2006; Collados Aís et al., 2007; Iglesias Fernández, 2013), since some are correlated constructs (Clifford, 2005; Yeh & Liu, 2008). The lack of consistent rating criteria (Collados Aís & García Becerra, 2015), however, precludes attempts at their operationalization, and, consequently, assessment in interpreting still lacks test reliability (Sawyer, 2004; Angelelli & Jacobson, 2009). Nevertheless, interpreting assessment has experienced great progress in terms of tools and resources. The use of rubrics, portfolios, reflective, deliberate and collaborative practice through technology-enhanced interpreting training platforms offers a myriad ways of interpreting practice (see, e.g., ORCIT, Speechpool, Interpreters in Brussels Practice Group), feedback (InterpretimeBank) and online training. However, the need still exists for a better understanding of the construct underlying the criteria as well as reliable measurements, which inform the design of tests, tools and resources used to assess students and provide them with feedback from trainers or their own peers.

Empirical research in interpreting testing and assessment grounded on testing theory (Angelelli & Jacobson, 2009) is being conducted to fill in this gap, with studies addressing interpreter competence and performance, and assessing processes and products for different purposes (i.e. training, industry). This allows for a more robust construct definition, operationalization of the notion of interpreting competence, skills and attributes (Skaaden & Wadensjö, 2014; Giamb Bruno, 2014; Hlavac & Orlando, 2015), aptitude, admission tests and screening (Bontempo & Napier 2009; Shlesinger & Pöchhacker 2011; Russo, 2011) as well as advances in standards and codes (Angelelli, 2006), and professional certification (Liu, 2013; Feuerle, 2013; Wallace, 2013; Hlavac & Orlando 2013; Han & Slatyer 2016). Different types of assessment formats, from rubrics (Angelelli, 2009; Jacobson, 2009) to portfolios (Sawyer 2004; Arumí Ribas, 2010) and different types of scales (Lee, 2008; Lee 2014, 2015; Tiselius 2009) are widely used in the interpreting classroom.

The benefits of formative assessment, in its various configurations: self-, peer and teacher assessment have been acknowledged, as it promotes students' better understanding of their strengths and weaknesses (Lee, 2005; Choi, 2006; Lee, 2016) and fosters metacognitive awareness (Witter-Merithew et al., 2001). Research on self-assessment in its various forms has proved to promote self-reflection and self-regulation (Russo 1995; Postigo Pinazo, 2008; Iaroslavschi, 2011, Sandrelli, 2015). Peer feedback is now provided in many interpreting courses (Lim, 2013; Lee, 2016). Instructors train students in conceptualizing assessment criteria to avoid the present lack of systematic criteria (Orsmond et al., 2010; Lee, 2016) while avoiding inter-rater variability in self- and peer assessment (Schjoldager, 1996; William & Thomson, 2007).

Equally, attempts have been made to avoid the lack of raters' understanding of scoring criteria (Collados Aís et al., 2007; Iglesias Fernández, 2006, 2013) and judgement consistency (Wu, 2013). In the realm of court interpreting, inroads into empirical research on testing models have been made by Wallace (2013), Giamb Bruno (2014), Hlavac and Orlando (2015), and Vigier Moreno and Valdés Garcés (2017), amongst others. Identifying competences and attributes for court interpreters has been at the heart of much research (the QUALITAS project; Giamb Bruno, 2014; Hlavac & Orlando, 2015).

Preface

Providing both translation and interpreting students with valuable feedback and implementing effective forms of assessment and practices are therefore essential not only for maximizing the teaching process, but also for enhancing students' learning experience. Translation/interpreting trainees expect information about industry assessment and revision practices and will need training to become future assessors themselves in their roles as revisers and reviewers, for instance (as provided in the European norm EN-15038, 2006, and in the new international standard ISO 17100, 2015). In other words, trainees need to practice how to observe translation/interpreting performances and translated/interpreted texts/discourses and how to tactfully communicate to a peer how the process or the end result could be improved (feedback). In addition, they need to be trained to assign a certain mark out of a scale to a translation/interpreting performance (assessment).

Observing, revising, giving feedback and assessing are issues where many of the debates on translation and interpreting training and practice intersect. This volume includes empirical contributions about competence assessment and quality, and the role of revision competence both in translation and interpreting training and in the industry. It presents ground-breaking methods of introducing the issue of translation quality assessment into training and reflects on innovative practices. Its findings explore and often support existing theoretical frameworks, but also point at refinements of the present scholarly work on translation quality assessment. However, in spite of the editors' efforts to include contributions directly from the industry, the latter unfortunately remain absent. Nevertheless, various chapters call out for collaboration between the industry and academia in order to enhance the validity of the theoretical constructs on the basis of empirical input from the professional domain.

AIMS AND TARGET AUDIENCE OF THE BOOK

This volume will serve as a guide for translation and interpreting researchers, academics, students and practitioners around the world to overcome the challenge of how translation and interpreting performances and results should be observed, given feedback to and assessed. By examining the changing role and function of revision, feedback and assessment, this book can inform the design of new ways of evaluating students as well as suggesting criteria for professional quality control. This ground-breaking contribution aims to provide a platform for researchers, scholars and professionals to incorporate into the classroom the changes that are presently affecting the translation and interpreting industries, such as new revision practices like post-editing and forms of translation (e.g., audiovisual translation and audiodescription), which lack academic training and call for further research. The descriptions of collaborative quality assurance practice initiatives, employing both general and revision/translation/interpreting-specific tools and resources, may complement professionals' goals with the innovative methods of feedback and assessment and turn their efforts into more comprehensive and effective sources of reflective and deliberate practice.

With its emerging approaches to Translation Quality Assessment, its discussions of how effective feedback can impact the acquisition of translation competence, its explorations of ways to incorporate Translation Quality Assessment into teaching through 21st-century methods such as peer feedback, and its connection of practices of revision to a feedback process in the industry by introducing students to professional quality standards, quality control criteria and benchmarks recognized at international level, this publication brings together ground-breaking areas of research. It provides comprehensive insight into up-to-date research on assessment practices in academic settings, and may thus inform training institutions of the way translation and interpreting trainees can be trained to become employable graduates. It

also offers trainees information about industry assessment practices. The recommendations offered in the book will address those translation educators with an interest in pedagogical research to introduce the new trends in assessment practices and feedback into their programmes, to enhance students' learning and to maximize teaching and learning methodologies and practices. This publication will notably contribute to the development of both fields of translation and interpreting, with a wide range of empirical case studies demonstrating innovation, experimental rigour and practical ideas and solutions to Translation and Interpreting scholars, educators and practitioners. The book also intends to play an essential role in proposing practical and empirically-based ways for universities and the industry to overcome traditional barriers to learning by promoting student and competence-centered training and effective ways to assess translation and interpreting quality.

ORGANIZATION OF THE BOOK

The book is divided into four sections and organized into 15 chapters. Section 1 (Chapter 1 to Chapter 3) provides an introduction to the field of translation quality assessment. Section 2 (Chapter 4 to Chapter 7) focuses on translation quality assessment in interpreting and audiovisual translation. Section 3 (Chapter 8 to Chapter 10) is devoted to process-oriented translation quality assessment, and Section 4 (Chapter 11 to Chapter 15) sheds some light into the learning process of students acquiring translation quality assessment competences. A brief description of each of the chapters follows:

Marcel Thelen introduces the reader to the domain of translation quality assessment with a survey of some conceptualizations of translation quality in both the translation industry, organizations and translator training programmes in Chapter 1. While various perspectives are described, this chapter mainly adopts the economic market-oriented perspective of translation as a provision of a service as outlined in the translation services standards EN 15038 and ISO 17100. He argues that translation quality assessment as it is performed in the industry should also be implemented in translation training curricula in such a way that students are made aware that translation as service provision also includes quality management and quality assurance. The ideal environment for such training is a so-called skills lab, he further suggests, explaining that it allows for the introduction of practicing both knowledge and skills that are essential in translation quality assurance and management. Applying the principles of validity and reliability to skills tests, the chapter further offers suggestions to translator educators for assessing students' activities in such skills labs in the areas of quality assurance, HR, client liaising, office management, team management and translation. He also argues that the last item - assessing the quality of the text-related translation activities - can both be performed top-down by translation trainers and bottom-up by students themselves.

Reliability and validity are also at the core of Chapter 2 by Gys-Walt van Egdom, Heidi Verplaetse, Iris Schrijver, Hendrik Kockaert, Winibert Segers, Jasper Pauwels, Bert Wylin and Henri Bloemen. While recognizing the complexities of the measurement of translation quality in all its ramifications, this chapter solely addresses the issue of quality evaluation of translation as a product. After introducing readers to translation evaluation methods, the authors illustrate and argue for the preselected items evaluation method (PIE method) by means of a case study. Taking recourse to the field of the exact sciences, they show that PIE has perturbative qualities, since it allows quality assessors to approximate the quality of a translation reliably and validly: preferably more than one assessor selects the items for evaluation, collects acceptable item translations and evaluates students' translations in more than one round, keeping a

Preface

critical eye on both the items selected and the translations accepted. Amongst other means, the authors put the test to the test by calculating the degree of difficulty of each item based on the students' results.

In Chapter 3, Gary Massey and Regine Weider approach translation quality assurance by exploring the unexploited potential that exists between translation and corporate communications. In the light of the rapid changes in the nature of translators' work as a result of technologization and other factors, the authors argue that the intercultural component intrinsic to translators could give them the opportunity to play an essential role in assuring quality in international corporate communications. By means of a pilot study at the interface between translation and corporate communications in Switzerland, the authors present and discuss the findings from a survey of translation and communications professionals. The chapter offers valuable insight into the current interplay between translation and corporate communications, and explores its implications for quality assurance and translator education.

Section 2 explores translation quality assessment in interpreting, dubbing and audiodescription settings. In Chapter 4, Chao Han argues that despite the increasing value attached to formative assessment, there is still the need for a longitudinally designed formative assessment model that harnesses the potential of self-, peer and teacher assessment. Han trialled this model in a case study in the consecutive English-Chinese interpreting classroom, involving self-, peer and trainer assessment. Grounded in testing theory, Han's model proposes a step by step approach to operationalizing a formative assessment model. The author further elicits students' and the teacher's perceptions of the assessment model. Based on the students' evaluations and the teacher's reflections, the chapter highlights potential contributions to effective formative assessment, discusses potential problems, proposes possible solutions, and suggests future trends in implementing and researching formative assessment in interpreter training.

In Chapter 5, Melissa Wallace questions the reliability and validity of the most extended oral assessment in interpreting certification tests for court interpreting in the United States and compares examinations conducted at state-level with those used in the federal courts. The oral exams are grounded in performance-based assessment and do not seem to draw from testing theory. The author proposes a model that brings together both a competence-based and a performance-based testing assessment model for this type of accreditation test. This requires a shift from the traditional single high-stakes examination to the assessment of a series of competence components involving skills and traits currently overlooked, such as soft skills and dispositional aptitudes, which are considered crucial for court interpreters.

Focusing on the field of audiovisual translation, Tomás Conde Ruano presents an approximation to the evaluation of quality in dubbing in Chapter 6. The author reports on the findings of a case study on how important different types of dischronies and the absence of image are for the evaluation of dubbing quality. Taking recourse to the field of translation expertise, the chapter discusses the differences in the data between two subgroups of subjects with different level of expertise (i.e. second-year undergraduate students and fourth-year undergraduate students). The results highlight how aspects such as the tone, the actors' elocution and the soundtrack may disguise possible translation or adjusting errors made when translating. The chapter concludes by addressing the implications of the results in audiovisual translation training.

Chapter 7 attempts to counteract the paucity of research into testing and assessment in audiodescription (AD), wherein Louise Fryer proposes a set of potential competences for AD student assessment with a view to guide the production of AD products. Acknowledging that AD users' priorities are not necessarily aligned with the competences materialized in AD scripts, her contribution draws on commonalities with simultaneous interpreting, as they share a common ground: They both are oral forms of translation of live events, and in spite of their own idiosyncrasies, many of the skills in AD and in

simultaneous interpreting overlap. The particularities of AD skills and competences are further discussed in the context of research by the ADLAB PRO Project. The author delves into assessment criteria in interpreting as compared to studies in AD with the aim to design an assessment sheet oriented to AD assessment to be used in AD quality assessment.

Section 3 of this volume is devoted to process-oriented translation quality assessment. While clients of translations are first and foremost interested in having quality translation products, translation service providers also need quality translation processes in order for their services to be profitable. Empirical studies of the translation process are often based on either Think Aloud Protocols, interviews, or other observational methods. In Chapter 8, Erik Angelone, however, explores the text-related processes by means of screen recording, which has already been used efficaciously to train translators. For assessment purposes, too, however, the protocols from screen recordings can be used by translation educators to reverse engineer the origins of errors found in translation products. Those protocols are argued to be useful for diagnosing triggers of errors in the translation process, because they contain observable indicators such as pauses, mouse hoverings, or deviations from established routines. They assist the assessor in identifying the locus of an error (whether it has been committed during comprehension, transfer or production activities), its phase (whether the error appeared at a drafting or a revision phase) and its information retrieval type (whether the translator accessed internal or external resources). Relying on screen recordings is further shown to increase inter-rater consistency and to enable the translation trainees themselves to self-assess their own processes, enhancing self-regulated learning.

In Chapter 9, Si Cheng introduces a problem-solving perspective to conceptualize the development of translation competence as part translation process assessment and sheds some light into the interrelation between the different subcompetences during the translation process. The translation problem-solving cycle proposed by the author involves the following steps: translation problem-identifying, translation problem-representing, translation solution-proposing, translation solution-evaluating, and translation decision-making sub-activities. The author's conceptualization is supported by an empirical longitudinal study conducted with translation trainees and involving the use of translation task-based interviews, two questionnaires and one focus group. The study reveals the need for process-oriented assessment to be able to gain information on students' progress and competence development, and suggests a more effective use of process-oriented pedagogical tools in translation education and training. Si Cheng's research thus offers not only a theoretical framework but also empirical evidence and practical suggestions to enhance process-oriented assessment.

The third contribution to process-oriented research of translation quality assessment focuses on revision behavior and applies both eye tracking and keylogging methods. In Chapter 10, Moritz Schaeffer, Anke Tardel, Sascha Hofmann and Silvia Hansen-Schirra study the eye movement and typing behavior during self-revision, which they define as the phase in the translation process which follows a first complete draft. The authors investigate the relative duration of the time that translators spend to revise the first complete draft of their source text and how that time is affected by various behavior characteristics during the drafting phase. Their study is based on a database of eye tracking and keylogging data from more than 300 translation sessions of 164 participants (about two thirds of students and one third of professionals) in total, translating English source texts into Spanish, Japanese, Danish, Hindi, Chinese or German. They obtain the result that the most efficient process involves a drafting phase with a large degree of concurrent reading and writing, and only few deletions. Although translators may go through a shorter revision stage if they avoid discontinuous typing, for example, the gains in total task time may be larger if they do the exact opposite.

Preface

While the last decade has seen a proliferation of assessment instruments which aimed at enhancing students' learning by prioritizing the development of translation competences and processes, it is yet unclear how translation trainees actually learn translation quality assessment. This aspect is dealt with in the book's final section. With their case study of a current MA Translation course, Elsa Huertas-Barros and Juliet Vine seek to provide a clearer insight into the current understandings of translation and assessment practices in Chapter 11. By taking a two-pronged approach, the authors survey the assessment practices in translation education, not only among the educators themselves but also among the students. Their contribution uses the framework of the six tenets of good assessment practice set out by the Higher Education Academy in the United Kingdom, including, for instance, the development of assessment fit for purpose, the recognition that assessment lacks precision and the need for a shared understanding of assessment and mutual trust. The authors also introduce the notion of assessment literacy and offer their study as an informed suggestion for a best practice of integrating assessment literacy into a course design.

Mari Pakkala-Weckström follows on from Elsa Huertas-Barros and Vine in encouraging students to become assessors and being involved in the assessment process. To this effect, Chapter 12 introduces a student self-evaluation grid providing a structured framework to evaluate both students' translation processes and products. Two different versions of the grid which were developed and tested with second- and third-year students are presented and supported by means of a pilot study. To enhance the proposed self-evaluation tool, students were asked to submit an end-of-course reflective commentary for each translation assignment, together with a portfolio including a reflective end-of-course self-evaluation. According to students' feedback, the enhanced version of the self-evaluation tool has proven to be a valuable instrument for self-monitoring and quality control, allowing students to conduct detailed retrospective analysis of their strengths and weaknesses.

Students' reflective and autonomous learning is also addressed by Ya-Yun Chen in Chapter 13. In the light of an empirical study conducted with MA students and involving both quantitative and qualitative methods, the author reports on students' level of reflection in different learning modes (i.e. use of diaries, think-aloud and small group discussion). A grounded theory-based qualitative analysis revealed that students generally adopted a reportive style in the diary and think-aloud modes, but a predominantly reflective style in small group discussions. The results also suggest that factors such as students' prior learning experience (including their cultural and educational background), their motivation and the characteristics, design and implementation of a reflective method should also be considered when implementing educational environments that promote students' reflective and autonomous learning.

In Chapter 14, Sonia Vandepitte and Joleen Hanson elaborate on the pedagogical effectiveness of collaborative projects in translation training, with a focus on the combination of revision with peer collaboration. It involves three studies addressing the effectiveness of peer comments and its relation to perceptions of expertise. It first discusses the role of translator's perceptions of reviewer expertise. It then inquires into the number and type of peer feedback comments in relation to whether the target language to be reviewed is the reviewer's L1 or L2 (native and non-native reviewers' feedback). In addition, translation direction is brought into the study by analysing the comments written by trainees in two situations: direct and inverse translation. Professional, methodological and theoretical research implications for translation and revision competence models are suggested.

The final chapter, Chapter 15, also sheds light on students' peer feedback performance: by means of questionnaires and focus groups, Carmen Heine investigates students' perceptions of feedback tasks and compares those with multiple peer feedback data. In fact, she also examines whether these methods can be operationalized as future peer activities in a translation didactics that implements a scaffolding

approach. The data derives from peer feedback related to a web-text localization and translation commentary task that was implemented in a double peer feedback loop that involved both students and teachers. The author analyzed both the wording of student feedback, the revisions suggested and the implementation by students of their peers' suggestions. She calls out for information about feedback and quality assurance practices to be provided by the profession, so that it can be incorporated into future student peer feedback practice, and enumerates various aspects of translation quality assessment that remain under-researched in the didactics of translation training.

Summarizing, the book presents a clear picture of some foremost principles and practices related to translation and interpreting quality assessment in the industry, translation education and translation studies today. It hopes to inspire colleagues in these areas to continue their efforts and collaborate further so as to enhance our insights into translation and interpreting practices and their reception in communities.

Elsa Huertas-Barros
University of Westminster, UK

Sonia Vandepitte
Ghent University, Belgium

Emilia Iglesias-Fernández
Universidad de Granada, Spain

REFERENCES

- Angelelli, C. (2006). Validating professional standards and codes: Challenges and opportunities. *Interpreting*, 8(2), 175–193. doi:10.1075/intp.8.2.04ang
- Angelelli, C. (2009). Using a rubric to assess translation ability: Defining the construct. In C. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies* (pp. 13–47). Amsterdam: John Benjamins. doi:10.1075/ata.xiv.03ang
- Angelelli, C. V., & Jacobson, H. E. (Eds.). (2009). *Testing and assessment in translation and interpreting studies. A call for dialogue between research and practice*. Amsterdam: John Benjamins.
- Arumí Ribas, M. (2010). Formative assessment in the interpreting classroom: Using the portfolio with students beginning simultaneous interpreting. *Current Trends in Translation Teaching and Learning*, 3, 97–133.
- Bontempo, K., & Napier, J. (2009). Getting it right from the start. Program admission testing of signed language interpreters. In C. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies* (pp. 246–295). Amsterdam: John Benjamins. doi:10.1075/ata.xiv.13bon
- Bühler, H. (1986). Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters. *Multilingua*, 5(4), 231–235.
- CEN. (2006). *Translation services: Service requirements, EN 15038*. Brussels: CEN.

Preface

- Chiaro, D., & Nocella, G. (2004). Interpreters' perception of linguistic and non-linguistic factors affecting quality: A survey through the World Wide Web. *Meta*, 49(2), 278–293. doi:10.7202/009351ar
- Choi, J. Y. (2006). Metacognitive evaluation method in consecutive interpretation for novice learners. *Meta*, 51(2), 273–283. doi:10.7202/013256ar
- Clifford, A. (2005). Putting the exam to the test: Psychometric validation and interpreter certification. *Interpreting*, 7(1), 97–131. doi:10.1075/intp.7.1.06cli
- Collados Aís, Á. (1998/2002). *La evaluación de la calidad en interpretación simultánea. La importancia de la comunicación no verbal*. Granada, Spain: Comares.
- Collados Aís, A., & García Becerra, O. (2015). Quality criteria. In F. Pöchhacker, N. Grbic, P. Mead, & R. Setton (Eds.), *Encyclopedia of Interpreting Studies* (pp. 368–383). London: Routledge.
- Collados Aís, Á., Iglesias Fernández, E., Pradas Macías, E. M., & Stévaux, E. (Eds.). (2011). *Qualitätssparameter beim Simultandolmetschen: Interdisziplinäre Perspektiven*. Tübingen: Narr Verlag.
- Collados Aís, Á., Pradas Macías, E. M., Stévaux, E., & García Becerra, O. (Eds.). (2007). *La evaluación de la calidad en interpretación simultánea: parámetros de incidencia*. Granada: Comares.
- Drugan, J. (2013). *Quality in professional translation assessment and improvement*. London: Bloomsbury.
- Feurerle, L. M. (2013). Testing interpreters: Developing, administering, and scoring court interpreter certification exams. *The International Journal of Translation and Interpreting Research*, 5(1), 297–323.
- Flanagan, M., & Heine, C. (2015). Peer-feedback as a translation training tool in Web-base communication. *Hermes*, 54, 115–136.
- Flanagan, M., & Heine, C. (2017). Peer-feedback as a translation training tool in web-based communication. *Journal of Language and Communication in Business*, 54, 115–136.
- Galán-Mañas, A., & Hurtado Albir, A. (2015). Competence assessment procedures in translator training. *The Interpreter and Translator Trainer*, 9(1), 63–82. doi:10.1080/1750399X.2015.1010358
- Giambruno, C. (2014). *Assessing legal interpreter quality through testing and certification: The QUALITAS project*. San Vicent del Raspeig, Spain: Publicaciones de la Universidad de Alicante.
- Gile, D. (1991). A communicative-oriented analysis of quality in nonliterary translation and interpretation. In M. L. Larson (Ed.), *Translation: Theory and practice, Tension and interdependence* (pp. 188–200). Binghamton, NY: SUNY. doi:10.1075/ata.v.19gil
- González Davies, M. (2004). *Multiple Voices in the Translation Classroom. Activities, Tasks and Projects*. Amsterdam: John Benjamins. doi:10.1075/btl.54
- Han, C., & Slatyer, H. (2016). Test validation in interpreter certification performance testing: An argument-based approach. *Interpreting*, 18(2), 231–258. doi:10.1075/intp.18.2.04han
- Hlavac, J., & Orlando, M. (2013). A cross-national overview of translator and interpreter certification procedures. *The International Journal of Translation and Interpreting Research*, 5(1), 32–65. doi:10.12807/ti.105201.2013.a02

- Hlavac, J., & Orlando, M. (2015). ‘Capturing’ and ‘prescribing’ desirable attributes of community interpreting: Design and implementation of testing systems in four anglophone countries. In C. Zwischenberger & M. Behr (Eds.), *Interpreting quality: A look around and ahead* (pp. 297–324). Berlin: Frank & Timme.
- Huertas Barros, E. (2011). Collaborative learning in the translation classroom: Preliminary survey results. *The Journal of Specialised Translation*, 16. Retrieved from http://www.jostrans.org/issue16/art_barros.php
- Huertas Barros, E. (2013). *La competencia interpersonal en la formación de traductores en España: Un estudio empírico-descriptivo del trabajo colaborativo durante la primera etapa de formación en Traducción e Interpretación* (Doctoral dissertation). Retrieved December 28, 2017, from <https://hera.ugr.es/tesisugr/22074508.pdf>
- Huertas Barros, E., & Vine, J. (2018). Current trends on MA translation courses in the UK: Changing assessment practices on core translation modules. *The Interpreter and Translator Trainer*, 12(1).
- Hurtado Albir, A. (1999/2003). *Enseñar a traducir: Metodología en la formación de traductores e intérpretes*. Madrid: Edelsa.
- Hurtado Albir, A. (2001/2004). *Traducción y traductología: Introducción a la traductología*. Madrid: Cátedra.
- Hurtado Albir, A. (2007). Competence-based curriculum design for training translators. *The Interpreter and Translator Trainer*, 1(2), 163–195. doi:10.1080/1750399X.2007.10798757
- Hurtado Albir, A. (2015). The acquisition of translation competence: Competences, tasks, and assessment in translator training. *Meta: Journal des Traducteurs / Meta: Translators’ Journal*, 60(2), 256–280.
- Iaroslavschi, M. (2011). Becoming an interpreter: Assessment and self-assessment in untutored practice sessions. *Cognition, Brain, Behavior: An Interdisciplinary Journal*, 15(2), 229–249.
- Iglesias Fernández, E. (2006). La indefinición del parámetro “agradabilidad de la voz” y los estudios de calidad de la interpretación simultánea. In M. J. Varela (Ed.), *La evaluación en los estudios de traducción e interpretación* (pp. 225–239). Seville, Spain: Bienza.
- Iglesias Fernández, E. (2013). Unpacking delivery criteria in interpreting quality assessment. In D. Tsagari & R. van Deemter (Eds.), *Assessment issues in language, translation and interpreting* (pp. 51–56). Frankfurt am Main: Peter Lang. Retrieved from <http://interpspracticegroup.com/about-us>
- InterpretimeBank. (n.d.). Retrieved from <http://interpretimebank.net>
- ISO 17100 (2015). *Translation services — Requirements for translation services*. Geneva: ISO.
- Jacobson, H. E. (2009). Moving beyond words in assessing mediated interaction. Measuring interactional competence in healthcare settings. In C. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies* (pp. 49–70). Amsterdam: John Benjamins. doi:10.1075/ata.xiv.04jac
- Kelly, D. (2005). *A handbook for translator trainers: Translation practices explained*. Manchester: St. Jerome.
- Kiraly, D. (2000). *A social constructivist approach to translator education: Empowerment from theory to practice*. Manchester: St Jerome Publishing.

Preface

- Klimkowski, K. (2006). Team work as a method in translation. *Across Languages and Cultures*, 7(1), 93–103. doi:10.1556/Acr.7.2006.1.6
- Kurz, I. (1993/2002). Conference interpretation: Expectations of different user groups. *The Interpreters' Newsletter*, 3, 13–21.
- Kurz, I., & Pöchhacker, F. (1995). Quality in TV interpreting. *Translatio-Nouvelles de la FIT – FIT Newsletter*, 14(3-4), 350-358.
- Lee, J. (2008). Rating scales for interpreting performance assessment. *The Interpreter and Translator Trainer*, 2(2), 165–184. doi:10.1080/1750399X.2008.10798772
- Lee, S.-B. (2014). An interpreting self-efficacy (ISE) scale for undergraduate students majoring in consecutive interpreting: Construction and preliminary validation. *The Interpreter and Translator Trainer*, 8(2), 183–203. doi:10.1080/1750399X.2014.929372
- Lee, S.-B. (2015). Developing an analytic scale for assessment undergraduate students' consecutive interpreting performances. *Interpreting*, 17(2), 226–254. doi:10.1075/intp.17.2.04lee
- Lee, S.-B. (2016). University students' experiences of 'scale-referenced' peer assessment for a consecutive interpreting examination. *Assessment & Evaluation in Higher Education*.
- Lee, Y.-H. (2005). Self-assessment as an autonomous learning tool in an interpretation classroom. *Meta*, 50(4). doi:10.7202/019869ar
- Lim, H.-O. (2013). A jury of your peers: Peer assessment in consecutive interpretation pedagogy. *Forum*, 11(1), 87–111. doi:10.1075/forum.11.1.04lim
- Lisaité, D., Vandepitte, S., Maylath, B., Mousten, B., Valdez, S., Castel-Branco, M., & Minacori, P. (2016). Negotiating meaning at a distance: Peer review in electronic learning translation environments. In M. Thelen, G. van Egdom, D. Verbeeck, Ł. Bogucki, & B. Lewandowska-Tomaszczyk (Eds.), *Translation and meaning: New series*, 41(1) (pp. 99–113). Frankfurt am Main: Peter Lang.
- Liu, M. (2013). Design and analysis of Taiwan's Interpretation Certification Examination. In D. Tsagari & R. van Deemter (Eds.), *Assessment issues in language translation and interpreting* (pp. 163-179). Frankfurt am Main: Peter Lang. Retrieved from <http://orcit.eu>
- Orsmond, P., Merry, S., & Reiling, K. (2010). The use of student-derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education*, 25(1), 23–38. doi:10.1080/02602930050025006
- Postigo Pinazo, E. (2008). Self-assessment in teaching interpreting. *TTR: traduction, terminologie, redaction*, 21(1), 173-211.
- Pradas Macías, E. M. (2006). Probing quality criteria in simultaneous interpreting: The role of silent pauses. *Interpreting*, 8(1), 25–43. doi:10.1075/intp.8.1.03pra
- Romero-Trillo, J. (2015). *Yearbook of corpus linguistics and pragmatics 2015: Current Approaches to discourse and translation studies*. Cham: Springer. doi:10.1007/978-3-319-17948-3
- Russo, M. (1995). Self-evaluation: The awareness of one's own difficulties as a training tool for simultaneous interpretation. *The Interpreters' Newsletter*, 6, 75–85.

- Russo, M. (2011). Aptitude testing over the years. *Interpreting*, 13(1), 5–30. doi:10.1075/intp.13.1.02rus
- Saldanha, G., & O'Brien, S. (2013). *Research methodologies in translation studies*. London: Routledge.
- Sandrelli, A. (2015). Becoming an interpreter: The role of computer technology. *MonTI*, 2, 111–138.
- Sawyer, D. (2004). *Fundamental aspects of interpreter education: Curriculum and assessment*. Amsterdam: John Benjamins. doi:10.1075/btl.47
- Schjoldager, A. (1996). Assessment of simultaneous interpreting. In C. Dollerup & V. Appel (Eds.), *Teaching translation and interpreting 3. New horizons: Papers from the third language international conference* (pp.187-195). Amsterdam: John Benjamins.
- Shlesinger, M., & Pöchhacker, F. (2011). Aptitude for interpreting. *Interpreting*, 13(1), 1–4. doi:10.1075/intp.13.1.01int
- Skaaden, H., & Wadensjö, C. (2014). Some considerations on the testing of interpreting skills. In C. Giambruno (Ed.), *Assessing legal interpreter quality through testing and certification: The QUALITAS project* (pp. 17-26). San Vicent del Raspeig, Spain: Publicaciones de la Universidad de Alicante. Speechpool. Retrieved from <http://www.speechpool.net/es/>
- Tiselius, E. (2009). Revisiting Carroll's scales. In C. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies* (pp. 95–121). Amsterdam: John Benjamins. doi:10.1075/ata.xiv.07tis
- Vandepitte, S., & Lisaité, D. (2016). *Expertise perceptions as a factor for effectiveness of peer feedback in translation training?* 8th EST Congress 'Translation Studies: Moving Boundaries', Aarhus, Denmark.
- Vigier Moreno, F. J., & Valdés Garcés, C. (2017). Shedding light on the grey area of assessment in legal translation: A case study based on a holistic method. *FITISPos International Journal*, 4, 11–27.
- Wallace, M. (2013). Rethinking bifurcated testing models in the court interpretation certification process. In D. Tsagari & R. van Deemter (Eds.), *Assessment issues in language translation and interpreting* (pp. 67–83). Frankfurt am Main: Peter Lang.
- Way, C. (2008). Systematic assessment of translator competence: In search of Achilles' heel. In J. Kearns (Ed.), *Translator and interpreter training: Issues, methods and debates* (pp. 88–103). London: Continuum International.
- Way, C. (2014). Structuring a legal translation course: A framework for decision-making in legal translator training. In L. Cheng, K. K. Sin, & A. Wagner (Eds.), *The Ashgate handbook of legal translation* (pp. 135–152). Farnham, UK: Ashgate.
- Way, C. (2016). Intra-university projects as a solution to the simulated/authentic dilemma. In D. Kiraly, L. Rüth, & M. Wiedmann (Eds.), *Towards authentic experiential learning in translator education* (pp. 147-159). Göttingen: V&R unipress/Mainz University Press.
- William, D., & Thomson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Mahwah, NJ: Erlbaum.

Preface

Witter-Merithew, A., Taylor, M., & Johnson, L. (2001). Guided self-assessment and professional development planning: A model applied to interpreters in educational settings. In C. Nettles (Ed.), *Proceedings of the 17th National Conference of the Registry of Interpreters for the Deaf* (pp. 153-226). Alexandria, VA: RID Publications.

Wu, F. S. (2013). How do we assess students in the interpreting examination? In D. Tsagari & R. van Deemter (Eds.), *Assessment issues in language translation and interpreting* (pp. 15–33). Frankfurt am Main: Peter Lang.

Yeh, S., & Liu, M. (2006). A more objective approach to interpretation evaluation: Exploring the use of scoring rubrics. *Journal of the National Institute for Compilations and Translations*, 34(4), 1–50.

Zwischenberger, C., & Pöchhacker, F. (2010). Survey on quality and role: Conference interpreters' expectations and self-perceptions. *Communicate (High Wycombe)*, 53. Retrieved from <http://www.aiic.net/ViewPage.cfm/article2510.htm>

Acknowledgment

The editors of this book are much indebted to many people who participated in this process. First of all, our thanks should go to all the contributors for their efforts and their professionalism in drafting the final versions of their chapters. The editorial process also much benefited from the expertise of our Editorial Advisory Board, for which we are very grateful. In particular, we are grateful to the contributors who also reviewed other chapters for this volume. They carried out a double task and showed impressive levels of quality and devotion: Ya-Yun Chen, Si Cheng, Tomás Conde Ruano, Chao Han, Silvia Hansen-Schirra, Carmen Heine, Sascha Hofmann, Moritz Schaeffer, Mari J. Pakkala-Weckström, Marcel Thelen, Gys-Walt van Egdom, and Melissa Wallace.

The idea for this book arose when we participated in the workshop entitled ‘New forms of feedback and assessment in translation and interpreting training and industry’ at the 8th EST Congress at Aarhus, in September 2016. We are fortunate to include some descriptions and reflections of studies that were presented, and we hope that participants at our workshop find this book stimulating.

Finally, we would like to thank IGI Global for their invitation to publish this book and especially, our Assistant Development Editor Maria, for her invaluable guidance and assistance throughout the process.

This book is dedicated to our families, who are always there: Miguel Ángel, Montserrat, Eva, Miriam, Preeti, Gert, Liisa, and Jana.

Section 1

Introduction to the Field of Translation Quality Assessment

Chapter 1

Quality and Quality Assessment in Translation: Paradigms in Perspective

Marcel Thelen

Maastricht School of Translation and Interpreting, The Netherlands

ABSTRACT

This chapter first gives an overview of existing interpretations of the concept of quality in organizations in general, as well as in the translation industry and translator training. This is followed by a discussion of quality assessment and how this is generally dealt with in the translation industry and translator training. Various perspectives are possible, but the focus is on translation as service provision as outlined in the translation services standards EN 15038 and ISO 17100. It is argued that implementing quality assessment of translation as service provision in translator training requires the introduction into the curriculum of quality management and quality assurance by students; a so-called skills lab is an ideal environment for this. Various suggestions are then given for assessing quality management and quality assurance while taking into account the requirements of validity and reliability.

INTRODUCTION

This chapter places existing paradigms of quality and quality assessment in perspective by discussing the various approaches to translation quality and its assessment in the translation industry and translator training. It also surveys the factors that play a role in quality assessment in translator training, and proposes a flexible assessment format. Translation is understood very broadly here, as any product or activity related to producing a target text on the basis of a source text. This can be translation *sec*, but also terminology lists, project management, office management related to translation, translation quality assurance, etc. Quality assessment can be carried out from the perspective of translation as a product, service, or process, but it can also focus on the translator, or the person carrying out the job in question. Central in this chapter will be translation as service provision. A good context for exploring this theme is a so-called in-house skills lab, or student company (or workplace) in the curriculum. Quality assessment of work done by students in such a skills lab involves more than just assessing translations and therefore also requires assessment methods, criteria and metrics other than those used for translation as product.

DOI: 10.4018/978-1-5225-5225-3.ch001

The chapter is structured as follows. First existing interpretations of quality are outlined, followed by a discussion of quality assessment in the translation industry and translator training. For translator training, quality assessment is then presented as the central concept that should measure the quality not only of translations and translation-related products, but also of student performance on quality management and quality assurance. The latter two form part of the service provision chain presented in the standards EN 15038 (CEN, 2006) and ISO 17100 (ISO 2015), and introduce service provision as a focal point of assessment. This is followed by a discussion of assessment methods, criteria and metrics in use and the factors influencing quality assessment, such as training method, assignment type, assessment method, criteria and metrics, competences (to be) assessed, etc. In order to take these factors into account in quality assessment, assessment grids are proposed that configure their interdependence and impact on assessment. A number of examples demonstrate how these grids may be used for a skills lab as well as for other training methods.

THE CONCEPT OF QUALITY

In Organisations in General

In organisations in general, the concept of quality is rather fuzzy since it may have various meanings. In his ISO 9000 Quality Systems Handbook, Hoyle (2001, p. 21), for example, lists the following possible meanings:

- A degree of excellence
- Conformance with requirements
- The totality of characteristics of any entity that bear on its ability to satisfy stated or implied needs
- Fitness for use
- Fitness for purpose
- Freedom from defects, imperfections or contamination
- Delighting customers

These meanings cover two perspectives: (1) inherent characteristics and (2) conformance with requirements, needs or expectations. The meaning “freedom from defects, imperfections or contamination” focuses on inherent characteristics, whereas the meaning “conformance with requirements” clearly points towards the perspective of conformance with requirements, needs or expectations. The other meanings seem to cover both perspectives.

It is perhaps for this reason that Hoyle (2001) proposes a combination of the two perspectives in an overall and uniform definition of quality as

... the degree to which a set of inherent characteristics fulfils a need or expectation that is stated, generally implied or obligator (Hoyle, 2001, p. 22).

where “stated” means stipulated in a document, and “generally implied” custom or common practice, whereas “obligatory” is not further explained and left to speak for itself (cf. ISO 9000:2005(E), p. 9).

“Characteristics” are further defined as

Quality and Quality Assessment in Translation

... inherent¹ characteristics ... of a product ..., process ..., or system ... related to a requirement ... (ISO, 2005, p. 12)

In this way, the definition relates to both the perspective of inherent characteristics and of conformance with requirements. Although not stated explicitly, the above two definitions (of quality and characteristics) also apply to (the provision of) services. Obviously, the characteristics of a product, process and system may well differ from one another, and, in the case of a product, the characteristics of e.g. a computer will differ from those of a candle, or a translation. In case the product is a translation, these characteristics will be linguistic (proper, appropriate and adequate language use in the target text) and the degree of the required type of equivalence between the source text and the target text.

In the Translation Industry

In the translation industry, the American Standard ASTM F2575-06², like ISO 9000:2005(E), combines “inherent characteristics” and “conformance” in one and the same definition of translation quality: “translation quality – the degree to which the characteristics of a translation fulfil the requirements of the agreed-upon specifications” (ASTM International, 2006, p. 4), and “...the requester’s expectations ...” (ASTM International, 2006, p. 10).

Koby, Fields, Hague, Lommel, and Melby (2014, pp. 4-5), on the other hand, give two definitions, viz. a broad one that explicitly refers to the perspective of “conformance” and a narrow one referring to the perspective of “inherent characteristics”. The broad definition reads:

A quality translation demonstrates accuracy and fluency required for the audience and purpose and complies with all other specifications negotiated between the requester and provider, taking into account end-user needs. (Koby et al., 2014, p. 416),

and the narrow one:

A high-quality translation is one in which the message embodied in the source text is transferred completely into the target text, including denotation, connotation, nuance, and style, and the target text is written in the target language using correct grammar and word order, to produce a culturally appropriate text that, in most cases, reads as if originally written by a native speaker of the target language for readers in the target culture. (Koby et al., 2014, pp. 416-417).

The specifications in the narrow definition refer to the degree of equivalence between source text and target text (which might be regarded as a form of conformity as well). In their view, the broad definition relates to translation quality management and the narrow one to translation quality; and choosing one instead of the other has an impact on the development of metrics for translation quality assessment. According to van der Meer et al.,

... [the narrow definition] might appeal to translators as it focuses on the linguistic quality of the target text, however, it does not address end-user needs, negotiated requirements or market adaptation and is therefore less applicable to commercial translation (van der Meer et al., 2017, p. 7)

In Translator Training

Whereas the translation industry remains implicit about combining product and service quality in one definition, translator training does combine the two perspectives: Gouadec (2013) defines quality as “both the quality of an end-product (the translated material) and the quality of the transaction (the service provided)” (Gouadec, 2013, p. 270). The first is called *intrinsic* and the second *extrinsic* quality (Gouadec, 2013, p. 272). Extrinsic quality is further defined as

(...) the way a translation satisfies the requirements of the applicable situation in terms of public (readers, viewers, browsers, listeners, etc.), objectives and purposes, medium or media, code, and such external parameters that are relevant (Gouadec, 2013, p. 272)

Gouadec proposes a number of quality grades for both types of quality, and argues that both should be included in quality assessment, proposing to make “quality-efficiency-relevance in service provision” and “quality of the end-product” interdependent (Gouadec, 2013, p. 275), thus combining the two perspectives for quality assessment.

Gouadec’s intrinsic quality, viz. “the quality of an end-product (the translated material)”, seems to follow the industry’s narrow definition (although not expressed explicitly) and includes source text-target text equivalence as well.

As is known, the relation between the source and target texts can be viewed from different angles and depends for a large part on the translation theory adhered to or applied and its view on translation quality. House (2015) distinguishes three main types of approach³: (1) psycho-social approaches, (2) response-based approaches, and (3) text and discourse-oriented approaches (House, 2015, pp. 8-20).

In the *psycho-social approach*, or mentalist view, the degree of translation quality depends on the impression the translation makes on its reader. Quality is assessed subjectively, impressionistically and intuitively by lay persons in terms of “how good or how bad one finds a translation” (House, 2015, p. 9) using such rather vague value statements as e.g. “The tone of the original is somehow lost in the translation” (House, 2015, p. 9). For this reason, this approach is not really suitable for objective quality assessment, although it might well be appropriate for assessing Newmark’s aspects of art and taste (see below).

Response-based approaches include approaches that are (1) behaviouristic or (2) functionalistic (i.e. “skopos”-related) (cf. House, 2015). In both, quality is viewed in terms of the nature of the reader’s response evoked by the translation. In the *behaviouristic approach*, quality is measured in terms of equivalence between the response from the readers of the source text and that from the readers of the target text. Nida and Taber (1969) termed this equivalence of response *dynamic equivalence*. Their assessment criteria were intelligibility and informativeness. The *functionalistic approach* based on Skopos theory (Vermeer, 1978; Nord, 1991, 1997) interprets quality as the degree to which the target text has the same intention, communicative purpose (“skopos”), (text) function and audience as the original source text. The instructions for the translator are laid down in the translation brief. In House’s view, Skopos theory

... is not very useful for translation quality assessment. (...) the notion of function is (...) never made appropriately explicit let alone operationalized (...). And exactly how the global skopos of a text is realized linguistically, and how one can determine whether a given translation fulfils its skopos, remains rather unclear (House, 2015, p. 11)

Quality and Quality Assessment in Translation

Text and discourse-oriented approaches comprise Descriptive Translation Studies (e.g. Toury, 1995), philosophical and socio-cultural, socio-political approaches (e.g. Venuti, 1995; Derrida, 1985 in Graham, 1985), and linguistically oriented approaches. The first mostly deal with literary translation. The other approaches are interested, on the one hand, in how a source text may be manipulated by the translator and may propagate the hidden motives of the original writer, and argue that the translator should be more visible (Venuti, 1995), or, on the other hand, in what meaning, text, and communication are (Derrida, 1985 in Graham, 1985). The main concern in text and discourse-oriented approaches is when a text is a translation.

In the context of translator training, the *linguistically oriented approaches* are perhaps the most interesting and most widely applied in professionally oriented curricula because these interpret equivalence between source text and target text in terms of linguistic phenomena. Well-known examples are Vinay and Darbelnet (1958), Catford (1965) and Chesterman (1997) with their translation strategies and procedures (known by others as translation techniques or transformations, cf. also Newmark, 1988; van Doorslaer, 2007).

Linguistically oriented approaches “provide detailed procedures for analysis and evaluation” (House, 2015, p. 14) and introduce all possible types of linguistic sub-disciplines to translation, such as corpus linguistics, discourse analysis, pragmatics, speech act theory, text linguistics, etc.

The purely individual or personal interpretation of translation quality remains, nevertheless, difficult to grasp. Newmark’s 1988 definition of translation is still relevant in this respect. He defines it as

1. *A science, which entails the knowledge and verification of the facts and the language that describes them – here, what is wrong, mistakes of truth can be identified*
2. *A skill, which calls for appropriate language and acceptable usage*
3. *An art, which distinguishes good from undistinguished writing and is the creative, the intuitive, sometimes the inspired, level of the translator*
4. *A matter of taste, where argument ceases, preferences are expressed, and the variety of meritorious translations is the reflection of individual differences (Newmark, 1988, p. 6)*

Newmark’s definition seems to relate more to the focal point of translation as a product than to other focal points (see the section *Focal Points of Assessment* below). It is in particular the last two parts of the definition (art and matter of taste) that point towards this individual or personal variation in interpretation of quality. These two aspects will undoubtedly have an impact on the reliability of translation quality assessment (see below).

Quality Management and Quality Assurance

Terms related to quality assessment are quality management and quality assurance. Both in the translation industry and in translator training there is disagreement about their use and they can also be found used interchangeably. The definition for quality management used in this chapter for translator training is based on ISO 9000:2005(E) and EN 15038/ISO 17100:

...coordinated activities to direct and control an organization... with regard to quality... (ISO 9000:2005(E), p. 9) that cover the whole process from first client-TSP⁴ contact to aftercare and complaints mechanism (EN 15038/ISO 17100)

Translation quality assurance is understood as

... a set of specific measures and procedures “... providing confidence that quality requirements will be fulfilled” (Hoyle, 2001, p. 59)

In this chapter, quality management pertains to the functioning and running of an organisation as a whole and incorporates quality assurance, whereas quality assurance focuses on the quality of a particular aspect of the organisation, e.g. translations as products in all their stages of creation.

Quality management and quality assurance are not yet a regular part of every curriculum, except in those institutes that have already introduced a skills lab⁵, student company⁶, or similar simulation of professional practice where both are done by students. Examples of such simulations are the in-house simulated translation bureau in Maastricht (NL) and the student company in Ghent (BE)⁷. Introducing quality management and quality assurance in the curriculum has an impact on quality assessment, viz. on the focal points of assessment, assessment methods and criteria; it also requires other types of training method and assignment and possibly also curriculum adjustments.

QUALITY ASSESSMENT⁸

In the Translation Industry

In the translation industry, generally two types of translation quality models can be distinguished: (1) top-down quality models and (2) bottom-up quality models (cf. Drugan, 2013, 2014). According to Drugan, this division into types occurs throughout the translation industry and applies to all forms of translation, from the “traditional” forms of translation such as translation, subtitling and so on to forms of collaborative translation⁹ such as crowd-sourcing. The two models differ in the way quality is achieved, managed and assured, and the impact this has on quality itself. A top-down quality model is developed and imposed by other agents than the person(s) actually carrying out a translation(-related) activity and comes, for example, in the form of a predetermined human or automated evaluation model with predetermined evaluation criteria, or quality standards (such as EN 15038 and ISO 17100). This type of quality model is rather rigid and “undemocratic” in that the actual performers of the task do not have any, or at least less, influence on how quality is (to be) achieved and assured. As Drugan (2014) puts it

Top-down approaches are hierarchical, driven from the top. They harness translation expertise and aim to manage or control quality levels. Associations with this group of approaches are traditional, conservative, authoritarian, or paternalistic (Drugan, 2014, p. 111)

A bottom-up translation quality model, on the other hand, is more flexible and gives performers/translators greater influence with or without the supervision and monitoring of a higher-order agent. According to Drugan (2014),

Bottom-up approaches, in contrast, are led by users or suppliers themselves. They draw on target-language ability and/or product expertise, combined with end-user feedback loops, rather than translation

Quality and Quality Assessment in Translation

competence. Associations with this group of approaches are novel, radical, egalitarian, or democratic (Drugan, 2014, p. 111)

As a result of this difference, the evaluation methods and criteria in bottom-up quality models are very likely to differ from those in top-down models because only the former allow performers/translators themselves to have a say in form and content of methods and criteria. In practice, however, very often a mix of both models is applied. For further features and examples of both models see Drugan (2013, 2014). In translator training, the most interesting element is who decides on how quality of student performance is (to be) achieved and assured.

The evaluation criteria and metrics in the translation industry for top-down evaluation are either human-operated, semi-automated or fully automated. They typically measure translation as a product and involve the identification of errors. Semi-automated evaluation takes place in localisation and the automotive sector, for example, where a threshold level of acceptability is set manually. Errors are spotted and tallied manually; the results are then compared automatically with the threshold level so as to calculate the quality of the translation in percentages of errors against error categories. An example of a semi-automated evaluation metric is the former LISA QA Model 3.1 (2006, for Localisation)¹⁰, and SAE J2450_201608 (for the automotive sector, SAE International, 2005). The error categories in a comparable metric, the J.D. Edwards' QA Form (a modification of the LISA QA Form) contain such error types as accuracy, style, grammar and formatting, each with its own weight (expressed in terms of percentages or using labels like "critical", "major" and "minor") and sub-categories:

- **Accuracy:** *Incorrect meaning, non-standard terminology, inconsistent terminology, untranslated SL;*
- **Style:** *Wrong register, inappropriate Anglicism, inappropriate use of passive/active voice;*
- **Grammar:** *Spelling errors, typos, grammar errors, syntax errors, punctuation errors;*
- **Formatting:** *Layout errors, font errors, double spaces (J.D. Edwards QA Form, n.d.)*

In the translation industry, especially in Machine Translation, translation quality is also measured in terms of the degree of correspondence of a translation product with reference translations (e.g. Phrase-based Statistical Machine Translation and Neural (Network) Machine Translation (see e.g. Koehn, 2017).

Although EN 15038 and ISO 17100 do not provide evaluation criteria and metrics, they list a number of requirements for translators to consider when translating which can be taken as evaluation criteria:

The translator shall translate in accordance with the purpose of the translation project, including the linguistic conventions of the target language and relevant project specifications. Throughout this process, the translator shall provide a service conforming to this International Standard with regards to the following:

1. *Compliance with specific domain and client terminology and/or any other reference material provided and ensuring terminological consistency during translation;*
2. *Semantic accuracy of the target language content;*
3. *Appropriate syntax, spelling, punctuation, diacritical marks, and other orthographical conventions of the target language;*
4. *Lexical cohesion and phraseology;*

5. *Compliance with any proprietary and/or client style-guide (including domain, register, and language variants);*
6. *Locale and any applicable standards;*
7. *Formatting;*
8. *Target audience and purpose of the target language content (ISO 17100, p. 10)*

A completely different approach is the Dynamic Quality Framework (DQF) developed by TAUS (Translation Automation User Society). The point of departure here is that the maxim that “translation users always need the highest quality” no longer holds and that “In the industry, there has been a recent focus on what constitutes acceptable levels of quality for different types of content and purposes” (van der Meer et al., 2017, p. 12). Elsewhere in the translation industry, this is called “fit for purpose” (cf. the DGT of the European Commission). The DQF uses three basic assessment criteria: Utility (“the relative importance of the functionality of translated content”), Time (“the time within which the translation is required”), and Sentiment (“the importance or impact on brand image, i.e. how potentially damaging might it be to a translation client if content is badly translated”) (cf. van der Meer, Görög, Dzeguze and Koot (2017, p. 12). These criteria function as “super” criteria for the choice of assessment model. Also in DQF, the focus seems to be on translation as product.

In Translator Training

Drugan’s distinction between top-down assessment and bottom-up assessment is highly relevant for translator training. If applied to translator training, a bottom-up quality model means that students, too, (to a large extent) decide on methods and criteria of quality assessment and how to use these. This may have consequences for the weight attached to such assessment in the training institute’s general assessment system of student performance. An additional practical issue may be how the newer types of translation that lend themselves to bottom-up quality assessment (such as crowd-sourcing) can be introduced into the curriculum.

However, in translator training, it is not (yet) common practice that student or bottom-up assessment has equal weight to trainer or top-down assessment nor that quality assurance and quality management by students are assessed and are regular parts of the curriculum. Introducing quality assurance and quality management by students requires a number of adjustments to existing practices and requires a new view on assessment, which will inevitably have an impact not only on quality assessment, but also on the curriculum.

This section discusses the most important factors that play a role in this respect such as (1) focal points of assessment, (2) assessment method, criteria and metrics, and (3) competences assessed. However, these are not the only factors that are involved. A discussion of the latter will conclude this section.

Focal Points of Assessment

In the translation industry, the focus of evaluation is generally on translation as a product, the provision of service (also called the translation process) and the translator. Product and translator are standard focal points in translator training and, increasingly, also service provision. In addition, training also focuses, or rather has to focus, on students’ learning needs and processes and the extent to which they have mastered the required competences and skills (this focal point could be called “education”). And

Quality and Quality Assessment in Translation

like the industry, training has to assess the results, but obviously with different consequences because of the completely different settings. In training – unlike the industry – continued bad results may lead to a delay in students completing their programme of study and in the worst case having to quit, although students always have the right to a second chance (Thelen, 2016a, p. 81).

Another focal point of assessment is the student translation process, not to be confused with the translation process as used in EN 15038 and ISO 17100, where it consists of translation, self-check, revision, review, proofreading and final verification and release (EN 15038, pp. 11-12 and ISO 17100, pp. 10-11). In translator training (and Translation Studies in general), translation process generally refers to the steps the translator takes to get from source language text to target language text, and includes the application of translation transformations/techniques/procedures, translation choices and decisions. To avoid confusion, from now on, the term translation cycle will be used for the sequence translation-self check-revision-review-proofreading-final verification-release.

When service provision is included as a focal point, it mostly concerns the delivery of the translation(-related) product as such, but not such aspects as quality management and quality assurance by students in line with EN 15038 and ISO 17100. Introducing these has a direct impact on quality assessment. It may be clear that the choice of focal points and how these are implemented will – partly – determine the methods, criteria and metrics for quality assessment.

Assessment Methods, Criteria and Metrics in Use

Most of the assessment methods used in translator training are top-down, i.e. mainly developed by staff against the background of Translation Studies, testing and assessment theories and practices, accreditation requirements, etc. Generally, these methods focus on the translation as a product (the target text) and/or the translation process (from source text to target text), in particular on translation errors. The translator, or rather the level of mastery of his or her competences, is also measured indirectly.

Overall, four aspects play a role in assessment methods: (1) who assesses, (2) the form of assessment, (3) how quality is assessed, and (4) the status of the assessment. The assessment can be done by staff, by visiting staff, by students or by persons outside of the institute such as placement hosts. Depending on the person who assesses, the status of the assessment is either formal (assessment by staff) or “advisory” (assessment by students or persons outside the institute); peer assessment is a form of “advisory” assessment¹¹. The assessment may be in the form of an exam (oral or written/on a computer), a portfolio, a translation project, an assessment interview (as in the case of a skills lab, for which see below), etc. There are a considerable number of assessment methods. Quality may be measured, for example, in full or using samples, analytically or holistically, or negatively or positively. Analytic quality assessment scrutinises a translation from beginning to end, mostly using linguistic criteria, whereas holistic assessment measures the quality of a translation more globally using such criteria as

- **Functional Adequacy (and Impact of Performance):** *The translation achieves its intended purpose and would achieve the intended effect on the reader.*
- **Quality of Content:** *The translation faithfully renders the arguments, reasoning and details of the source text (coherence and attention to detail).*
- **Textual Adequacy:** *The translation is idiomatic, readable and typographically and idiomatically correct.*

- **Quality of Specialized Content:** *The translator uses terminology from authoritative, relevant sources and specialized concepts are faithfully conveyed (Colina, 2009, 259-260).*

Alternatively, samples of a translation can be assessed, as in the Preselected Items Evaluation (PIE) method, (e.g. Segers & Kockaert, 2016), and the Calibration of Dichotomous Items (CDI) method (Eyckmans, Anckaert, & Segers, 2009); for a comparison of the two, see Kockaert and Segers (2017). Typical of these methods is that translation is assessed mainly as a product. They have in common that they compare a translation with some kind of standard in a dichotomous way: in terms of correctness or incorrectness. In the case of PIE, there are a number of preselected items; performance on these items is scored and constitutes the ultimate quality verdict; the selection of items is done by expert assessors. In the case of CDI, preselected items also form the basis of quality measurement, but their selection is different: those items are selected that in a pre-test translation are translated correctly by translation students with a good translation competence, and incorrectly by students with a poor translation competence. Both methods claim reliability of their assessments.

Characteristic of linguistically oriented approaches is that they provide detailed lists of assessment criteria. Al-Qinai (2000), for example, uses the following criteria:

1. **Textual Typology (Province) and Tenor:** *i.e. the linguistic and narrative structure of ST and TT, textual function (e.g. didactic, informative, instructional, persuasive, evocative... etc.).*
2. **Formal Correspondence:** *Overall textual volume and arrangement, paragraph division, punctuation, reproduction of headings, quotation, mot[t]os, logos... etc.*
3. **Coherence of Thematic Structure:** *Degree of referential compatibility and thematic symmetry.*
4. **Cohesion:** *Reference (co-reference, proforms, anaphora, cataphora), substitution, ellipsis, deixis and conjunctions.*
5. **Text-Pragmatic (Dynamic) Equivalence:** *Degree of proximity of TT to the intended effect of ST (i.e. fulfilment or violation of reader expectations) and the illocutionary function of ST and TT.*
6. **Lexical Properties (Register):** *Jargon, idioms, loanwords, catch phrases, collocations, paraphrases, connotations and emotive aspects of lexical meaning.*
7. **Grammatical/ Syntactic Equivalence:** *Word order, sentence structure, cleaving, number, gender and person (agreement), modality, tense and aspect (Al-Qinai, 2000, p. 499)*

These criteria are further divided into various sub-criteria. Implicit in most assessment methods, criteria and metrics in translator training is that quality should always be as high as possible, irrespective of the type and purpose of source text and of the target text. In this respect, the Dynamic Quality Framework of TAUS could be an interesting asset in translator training.

Factors Influencing Quality Assessment in Translator Training

The factors that may have an impact on assessment include the type of training method (e.g. individual work, group work, working in a skills lab, etc.), type of translation(-related) assignment (e.g. translating, managing a skills lab, setting up a quality assurance system, terminology work, etc.), mode of translation (e.g. translation, subtitling, localisation, etc.), specialist domain (e.g. law, medicine, economics, etc.) as well as the type of translation (e.g. traditional translation, oral or written, plain or annotated, on-sight translation, collaborative translation such as in crowd-sourcing, etc.). In addition, the various focal points

Quality and Quality Assessment in Translation

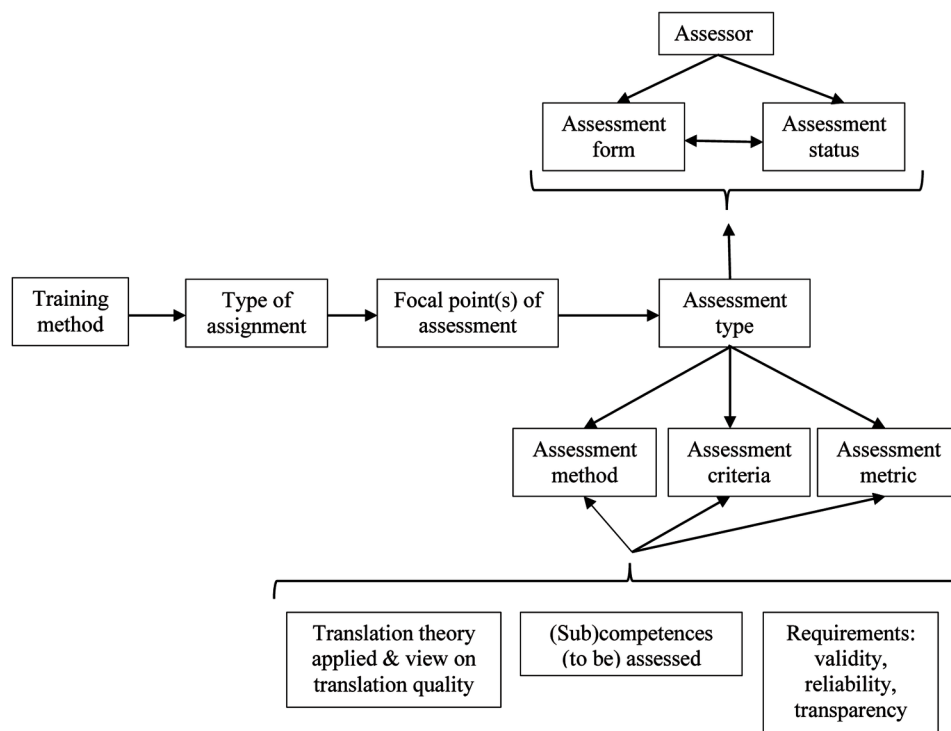
do not assess all and the same competences, nor is the type of feedback to students the same in all cases. In other words, there is a great interdependence between all these factors.

The interrelationship between the various factors that have an impact on assessment is visualised in Figure 1.

Legend to Figure 1:

- **Assessment Form:** An examination (oral or written/on a computer), a portfolio, a translation project, coach sessions, an assessment interview with management of a skills lab (see below), assessment interview with the trainer-director of a skills lab (see below), documents produced (e.g. by the quality assurance officer in a skills lab), feedback from clients and skills lab staff (= students), etc.
- **Assessment Method:** By samples; in full; holistically; positive/negative remarks in feedback from clients/skills lab staff, or adequate points in quality assurance document (positive assessment); by counting translation mistakes and errors, etc. (negative assessment), etc.
- **Assessment Type:** Top-down (by trainer or clients/customers) or bottom-up assessment (by a skills lab's management or HR officer).
- **Translation Theory Applied and View on Quality:** Psycho-social approaches, response-based approaches, text and discourse oriented approaches, etc.
- **Type of Translation(-Related) Assignment:** Translating, managing a skills lab, acting as a quality assurance officer, terminology work, etc.; “traditional” translation – oral or written, plain or annotated –, on-sight translation, collaborative translation such as crowd-sourcing, etc.

Figure 1. Factors influencing quality assessment in translator training



- **Type of Training Method:** Skills lab, individual translation work, group translation work

Figure 1 is to be interpreted as follows. There is a relation of dependency between the type of training method and the type of assignment: a skills lab, for example, allows/calls for other types of assignment than just translating, such as quality management, quality assurance, client contact, running a business, etc. The dependency is two-way: a training method may determine the type of assignment, and the type of assignment may require a particular training method. Both have an impact on the focal point of quality assessment: in the case of a translation (the assignment type) in individual translation work (the training method), the focal points of quality assessment are either product, process, translator, or all of these; in the case of liaising with clients (the assignment) in a skills lab (the training method), the focal point of quality assessment is obviously translation as service provision.

The assessment type can be either top-down or bottom-up. Liaising with clients in a skills lab with service provision as the focal point of assessment, for example, will be assessed top-down with clients and the trainer-director of the skills lab as assessors, and bottom-up with the skills lab management (i.e. the student charged with HR) as assessor. In the case of clients as assessors, the assessment could be “advisory”; in the case of the trainer-director of the skills lab as assessor, the assessment could be formal. In the case of the HR officer in the skills lab management (who is a peer) as the assessor, the assessment is peer assessment and is most likely “advisory”. The assessment type will determine the assessment methods, criteria and metrics (if applicable): in cases where the skills lab’s HR officer assesses client liaison (peer assessment, bottom-up assessment), for example, the assessment method will take into account the nature of the remarks by clients and perhaps also other staff in the skills lab (e.g. complaints by translators); the assessment criteria will then be positive and negative remarks (viz. compliments and complaints); for time reasons there will probably be no detailed metric, but the balance between compliments and complaints will determine the final “advisory” assessment. In the case of top-down assessment of a translation (by the trainer-director of the skills lab), for example, this picture is different. Then, the particular translation theory applied and its view on translation quality, thus the relationship between source text and target text, will determine the assessment criteria used (e.g. linguistically oriented or not), and will, consequently, have an impact on the assessment metric.

At the same time, the translation as assignment also determines the focal points of assessment. In all cases, the competences (to be) assessed and the requirements of validity, reliability and transparency play an important role and have to be accounted for vis-à-vis exam board and accreditations. Transparency is a typical training requirement and stipulates that students should know precisely what it is that they are set to do, what learning objectives are involved, by what criteria they will be assessed and that they will get feedback on their results. Validity is defined by Williams (2009) as “the extent to which an evaluation measures what it is designed to measure”, which he calls “construct validity” (Williams, 2009, p. 5). In addition, he distinguishes two other types of validity: “content validity” (“the extent to which an evaluation covers the skills necessary for performance”, Williams, 2009, p. 5), and “predictive validity” (“[the extent in which the evaluation] accurately predict[s] future performance”, Williams, 2009, p. 5). He defines reliability as “the extent to which an evaluation produces the same results when administered repeatedly to the same population under the same conditions.” (Williams, 2009, p. 5); reliability includes reproducibility and objectivity.

Not all competences will be assessed/assessable by every training method, type of assignment, assessment type, assessment form, and assessment method, nor will they be suitable for every focal point of assessment. Finally, type, mode and specialist domain of translation are also factors of influence:

Quality and Quality Assessment in Translation

collaborative translation (e.g. in crowd-sourcing) behaves differently than on-sight translation. Subtitling, for example, requires other assessment methods, criteria and metrics than localisation; and the view on translation quality (i.e. the conformance between source text and target text) for legal translations differs from that for medical translations. The interrelationship between the above factors shows that assessment is not an easy task and should be dealt with very carefully so as to achieve the optimal result.

Competences Assessed

Translation quality assessment is not a stand-alone procedure in the process of judging and measuring quality in translation, but is linked very closely to other elements, for example (sub-)competences to be assessed. Assessment cannot not always assess all or the same (sub-)competences formulated. This is due to the type of assessment, but also to the assignment given: not every assignment is suited for every type of assessment (Angelelli, 2009, p. 37).

Among the sets of competences available for translator training, those formulated by the PACTE Group (Hurtado Albir, 2017), the EMT competences (EMT expert group, 2009 - a revised version is soon to be made public), and the EN 15038/ISO 17100 (CEN, 2006) are the most well-known. The EMT competences also explicitly include translation service provision, whereas the other sets mostly relate to the production of a translation, and focus on target text, process and translator. The EMT translation service provision competence seems the only competence that is able to play a role in quality management and assurance by students (and thus in a skills lab) as its components are very explicit. The EMT translation service provision competence has an interpersonal dimension and a production dimension, both with various sub-competences:

- *Interpersonal Dimension:*
 - *Being aware of the social role of the translator*
 - *Knowing how to follow market requirements and job profiles (knowing how to remain aware of developments in demand)*
 - *Knowing how to organise approaches to clients/potential clients (marketing)*
 - *Knowing how to negotiate with the client (to define deadlines, tariffs/invoicing, working conditions, access to information, contract, rights, responsibilities, translation specifications, tender specifications, etc.)*
 - *Knowing how to clarify the requirements, objectives and purposes of the client, recipients of the translation and other stakeholders*
 - *Knowing how to plan and manage one's time, stress, work, budget and ongoing training (upgrading various competences)*
 - *Knowing how to specify and calculate the services offered and their added value*
 - *Knowing how to comply with instructions, deadlines, commitments, interpersonal competences, team organisation*
 - *Knowing the standards applicable to the provision of a translation service*
 - *Knowing how to comply with professional ethics*
 - *Knowing how to work under pressure and with other experts, with a project head (capabilities for making contacts, for cooperation and collaboration), including in a multilingual situation*
 - *Knowing how to work in a team, including a virtual team*

- *Knowing how to self-evaluate (questioning one's habits; being open to innovations; being concerned with quality; being ready to adapt to new situations/conditions) and take responsibility*
- *Production Dimension:*
 - *Knowing how to create and offer a translation appropriate to the client's request, i.e. to the aim/skopos and to the translation situation*
 - *Knowing how to define stages and strategies for the translation of a document*
 - *Knowing how to define and evaluate translation problems and find appropriate solutions*
 - *Knowing how to justify one's translation choices and decisions*
 - *Mastering the appropriate metalanguage (to talk about one's work, strategies and decisions)*
 - *Knowing how to proofread and revise a translation*
 - *(mastering techniques and strategies for proofreading and revision)*
 - *Knowing how to establish and monitor quality standards (EMT expert group, 2009, p. 4-5)*

Most of these sub-competences are important in a skills lab and clearly relate to the focal point of assessment of service provision. In particular, this applies to most of the sub-components of the interpersonal dimension and the last component of the production dimension.

QUALITY ASSESSMENT IN A SKILLS LAB

In comparison to other training methods, a skills lab, or similar training method, opens up many more opportunities for variation in assessment and is more in line with the translation industry as it makes the introduction of quality management and assurance by students possible. It is, moreover, a valuable contribution to minimising the alleged skills gap between training and industry, and may help to make the transition as smooth as possible. Because a skills lab allows for a great variety of tasks for students, quality assessment cannot be uniform. Moreover, since students are individually responsible and accountable for some tasks (e.g. translating, quality assurance, etc.) while they are responsible and accountable for other tasks as a group (e.g. running the simulated translation bureau, making profit, etc.), two types of assessment grid are required: one for individual assessment and one for group assessment. The following grids may be helpful to capture these variations in quality assessment and the above-mentioned factors influencing them. The exact details depend on the local situation. Mode and specialist domain of translation may be specified as details of translation assignment. The following tables are examples only. Table 1 represents a grid for top-down assessment of quality assurance in a skills lab.

The grid for quality management, in particular the function of HR officer, would be as shown in Table 2.

In this way, more grids can be composed for other functions in a skills lab. Table 3 represents top-down assessment of client liaison in a skills lab.

Top-down assessment of office management is presented in Table 4.

Table 5 represents a grid for top-down assessment of translating in a skills lab¹³.

Students may produce a large number of translations in a number of different language combinations. In the Maastricht skills lab set-up it is hardly possible for the trainer-director to assess each and every translation analytically. Similarly, the clients (who are trainers mostly) also get great numbers of translations to assess. To facilitate fast assessment of each translation, the assessment metric was kept rather simple (scores from 1 to 5 on a set of rather general assessment criteria); consequently, assessments are

Quality and Quality Assessment in Translation

Table 1. Assessment grid for top-down assessment of quality assurance in a skills lab

Factor	Details
Training method:	Skills lab
Type of assignment:	Quality assurance, function/task: quality officer
Focal point of assessment:	Translation as service provision
Assessment type:	Top-down
Assessor:	Trainer-director of skills lab
Assessment status:	Formal
Assessment form:	Judging and measuring the quality of documents produced, functioning as quality officer, the nature of the feedback given by clients, feedback given by skills lab staff, ...
Assessment method:	By samples, holistically, and by positive and negative assessment
Assessment criteria:	<i>Documents produced:</i> are they complete and according to instructions, correct, coherent and adequate, is the system well-documented, ...? <i>Functioning:</i> were there any problems, were problems adequately solved, ...? <i>Management records of client feedback:</i> positive, negative, any compliments, any complaints, ...? <i>Management records of feedback from skills lab staff:</i> positive, negative, any compliments, any complaints, ...?
Assessment metric:	All in terms of YES/NO; the balance between them determines the final score, if YES, then PASS, if NO, then FAIL
Translation theory applied:	n.a.
Competences assessed ¹² :	Translation service provision competence (EMT), linguistic and textual competence (EN 15038/ISO 17100), competence in research, information acquisition and processing (ISO 17100), technical competence (EN 15038/ISO 17100)
Assessment requirements:	Validity: the assessment measures quality assurance and nothing else, covers the skills necessary for quality assurance and predicts adequate future performance accurately
	Reliability: because of the dichotomous character of the scoring options on the metric, the assessment is reproducible and objective, thus reliable
	Transparency: students are given clear instructions beforehand on the assessment and receive various forms of feedback.

Table 2. Assessment grid for top-down assessment of HR in a skills lab

Factor	Details
Training method:	Same as in Table 1
Type of assignment:	Management, function/task: HR officer
Focal point of assessment:	Same as in Table 1
Assessment type:	Same as in Table 1
Assessor:	Same as in Table 1
Assessment status:	Same as in Table 1
Assessment form:	Judging and measuring the quality of documents produced, functioning as HR officer, the nature of the feedback given by clients, feedback given by skills lab staff, ...
Assessment method:	Same as in Table 1
Assessment criteria:	<i>Documents produced:</i> e.g. reports of assessment interviews with skills lab staff: are they complete and according to instructions, correct, coherent and adequate, is the system well-documented, ...? <i>Functioning:</i> were there any problems, were problems adequately solved, ...? <i>Management records of feedback from skills lab staff:</i> positive, negative, any compliments, any complaints, ...?
Assessment metric:	Same as in Table 1
Translation theory applied:	Same as in Table 1
Competences assessed:	Same as in Table 1
Assessment requirements:	Validity: the assessment measures HR and nothing else, covers the skills necessary for HR and predicts adequate future performance accurately
	Same as in Table 1
	Same as in Table 1

Table 3. Assessment grid for top-down assessment of client liaising in a skills lab

Factor	Details
Training method:	Same as in Table 1
Type of assignment:	Management, function/task: client liaison officer
Focal point of assessment:	Same as in Table 1
Assessment type:	Same as in Table 1
Assessor:	Same as in Table 1
Assessment status:	Same as in Table 1
Assessment form:	Judging and measuring the quality of documents produced, functioning as client liaison officer, the nature of the feedback given by clients, feedback given by skills lab staff, ...
Assessment method:	Same as in Table 1
Assessment criteria:	<i>Documents produced:</i> e.g. records of client contacts arrangements made, client instructions/requirements, translation briefs, complaints and complaint handling, etc. are they complete and according to instructions, correct, coherent and adequate, is the system well-documented, ...? <i>Functioning:</i> were there any problems, were problems adequately solved, complaints handled to the satisfaction of clients...? <i>Management records of feedback from skills lab staff:</i> positive, negative, any compliments, any complaints, ...?
Assessment metric:	Same as in Table 1
Translation theory applied:	Same as in Table 1
Competences assessed:	Same as in Table 1
Assessment requirements:	Validity: the assessment measures client liaising and nothing else, covers the skills necessary for client liaising and predicts adequate future performance accurately
	Same as in Table 1
	Same as in Table 1

Table 4. Assessment grid for top-down assessment of office management in a skills lab

Factor	Details
Training method:	Same as in Table 1
Type of assignment:	Management, function/task: office manager
Focal point of assessment:	Same as in Table 1
Assessment type:	Same as in Table 1
Assessor:	Same as in Table 1
Assessment status:	Same as in Table 1
Assessment form:	Judging and measuring functioning as office manager, the nature of the feedback given by clients, feedback given by skills lab staff, ...
Assessment method:	Same as in Table 1
Assessment criteria:	<i>Functioning:</i> leadership, management, steering and motivating staff: were there any problems, were problems adequately solved, complaints handled to the satisfaction of clients...? <i>Records of client feedback:</i> positive, negative, any compliments, any complaints, ...? <i>Management records of feedback from skills lab staff:</i> positive, negative, any compliments, any complaints, ...?
Assessment metric:	Same as in Table 1
Translation theory applied:	Same as in Table 1
Competences assessed:	Same as in Table 1
Assessment requirements:	Validity: the assessment measures office management and nothing else, covers the skills necessary for office management and predicts adequate future performance accurately
	Same as in Table 1
	Same as in Table 1

Quality and Quality Assessment in Translation

Table 5. Assessment grid for top-down assessment of translating in a skills lab

Factor	Details
Training method:	Same as in Table 1
Type of assignment:	Translating, function/task: translator
Focal point of assessment:	Translation as a product, translation process (from source text to target text), translation cycle (whole sequence from translation, self-check, revision, review, proofreading and final verification and release), translator
Type of translation	Written "traditional" translation with CAT tools
Assessment type:	Same as in Table 1
Assessor:	Same as in Table 1
Assessment status:	Same as in Table 1
Assessment form:	Judging and measuring the quantity and quality of translations produced, functioning as a translator, the nature of the feedback given by clients, feedback given by skills lab staff, ...
Assessment method:	Same as in Table 1
Assessment criteria:	<i>Quantity of translations produced</i> : has the required number of words been translated? <i>Quality of the translation produced</i> : spelling, grammar, style, word choice, equivalence between source and target text, conformance with instructions, style guide, glossary <i>Functioning in skills lab</i> : were there any problems, were problems adequately solved, complaints handled to the satisfaction of clients...? <i>Management records of client feedback</i> : positive, negative, any compliments, any complaints, ...? <i>Management records of feedback from skills lab staff</i> : positive, negative, any compliments, any complaints, ...?
Assessment metric:	<i>Translations produced – quantity of translations</i> : in terms of YES/NO; the balance between them determines the final score, if YES, then PASS, if NO, then FAIL <i>Functioning in skills lab, management records of client feedback, management records of feedback from skills lab staff</i> : all in terms of YES/NO; the balance between them determines the final score of translations: if YES, then PASS, if NO, then FAIL <i>Translations produced – quality of translations</i> : scores from 1 (very poor, unsatisfactory) to 5 (excellent)
Translation theory applied:	Only applicable to <i>translations produced – quality of translations</i> : depends on local situation. Mostly linguistically oriented approach
Competences assessed:	Translation service provision competence (EMT), linguistic and textual competence (EN 15038/ISO 17100), competence in research, information acquisition and processing (ISO 17100), technical competence (EN 15038/ISO 17100), intercultural competence (EMT)
Assessment requirements:	Validity: the assessment measures translating and nothing else, covers the skills necessary for translating and predicts adequate future performance accurately
	Reliability: for all the assessment forms and criteria except for <i>translations produced – quality of translations</i> : because of the dichotomous character of the scoring options on the metric, the assessment is reproducible and objective, thus reliable For <i>translations produced – quality of translations</i> : reliability cannot be guaranteed
	Same as in Table 1

perhaps not thorough enough and not fully reliable. However, for his overall assessment of translation performance of the individual skills lab translators, the trainer-director assesses a number of samples from a selection of translations of each translator, and combines this assessment with the “advisory” assessments from clients and skills lab revisers. This increases the degree of objectivity of the final overall assessment somewhat, although there is no guarantee that the final score will be the same if and when the assessment is repeated. Clearly, this needs improvement.

In the case of bottom-up (peer) assessment as presented in Table 6, no guarantee can be given for reliability, at least not for individual translations. Although students receive clear and strict instructions for assessing, personal and subjective motives may influence their assessment. This effect is undone somewhat when more students (revisers and management in verification stage before delivery) assess more translations by the same translator; then a reasonably valid and reliable picture may emerge. Validity cannot be guaranteed either, since students may unconsciously allow other factors to play a role in their

Table 6. Assessment grid for bottom-up assessment of translating in a skills lab

Factor	Details
Training method:	Skills lab
Type of assignment:	translating, function/task: translator
Type of translation	Written “traditional” translation with CAT tools
Focal point of assessment:	Translation as a product, translation process (from source text to target text), translation cycle (whole sequence from translation, self-check, revision, review, proofreading and final verification and release), translator
Assessment type:	Bottom-up
Assessor:	Management (in final verification stage before delivery), revisers (in translation cycle)
Assessment status:	“advisory”, peer
Assessment form:	Judging and measuring the quantity and quality of translations produced, functioning as a translator, ...
Assessment method:	Same as in Table 1
Assessment criteria:	<i>Translations produced – quality of translations:</i> spelling, grammar, style, word choice, degree of equivalence between source and target text, conformance with client’s instructions, style guide, glossary, ... <i>Functioning in skills lab:</i> keeping deadlines, cooperation, were there any problems, were problems adequately solved/complaints handled to the satisfaction of clients...?
Assessment metric:	scores from 1 (very poor, unsatisfactory) to 5 (excellent)
Translation theory applied:	Only applicable to <i>translations produced – quality of translations:</i> depends on local situation. Mostly linguistically oriented approach
Competences assessed:	Translation service provision competence (EMT), linguistic and textual competence (EN 15038/ISO 17100), technical competence (EN 15038/ISO 17100), intercultural competence (EMT)
Assessment requirements:	Validity: cannot be guaranteed
	Reliability: cannot be guaranteed
	Same as in Table 1

Table 7. Assessment grid for top-down assessment of group work (management) in a skills lab

Factor	Details
Training method:	Same as in Table 1
Type of assignment:	Successfully running a simulated translation bureau: management as a group
Focal point of assessment:	Same as in Table 1
Assessment type:	Same as in Table 1
Assessor:	Same as in Table 1
Assessment status:	Same as in Table 1
Assessment form:	Same as in Table 1
Assessment method:	Same as in Table 1
Assessment criteria:	<i>Documents produced:</i> e.g. order portfolio, workflow management system, accounts/books, client satisfaction system/complaints system, etc. are they complete and according to instructions, correct, coherent and adequate, is the system well-documented, has a profit been made...? <i>Functioning:</i> in terms of day-to-day running of the business: were there any problems, were problems adequately solved, complaints handled to the satisfaction of clients...? <i>Feedback from clients:</i> positive, negative, any compliments, any complaints, ...? <i>Feedback from skills lab staff:</i> positive, negative, any compliments, any complaints, ...?
Assessment metric:	Same as in Table 1
Translation theory applied:	Same as in Table 1
Competences assessed:	Same as in Table 1
Assessment requirements:	Validity: the assessment measures managing a simulated translation bureau and nothing else, covers the skills necessary for managing a translation bureau and predicts adequate future performance accurately
	Same as in Table 1
	Same as in Table 1

Quality and Quality Assessment in Translation

assessment(s) such as affection, antipathy, etc. Clearly, reliability and validity of peer assessments need improvement, all the more so because peer assessment is a very useful learning mechanism for students (Li, Steckelberg, & Srinivasan, 2008). Introducing newer types of translation such as crowd-sourcing in the curriculum could be an asset to translation quality assessment and contribute to the further implementation of bottom-up assessment (peer assessment).

Tables 1-6 pertain to individual work and presuppose a particular set-up of the skills lab with the functions/tasks indicated. The functions/tasks may, however, also be combined in fewer individual students depending on the total number of students taking part in the skills lab. There should, after all, be a healthy balance between management and translators/revisers. Table 7 presents an assessment grid for group work.

Here, too, the assessment metric in terms of YES (PASS) – NO (FAIL) provides a good basis for validity and reliability.

The grid model can also be used for other training methods such as individual translation work. In individual translation work (e.g. as an examination) fewer details (in the sense of the Tables above) are involved. In spite of this, however, the assessment criteria and metric are more complicated and assessment may require more time and energy. Table 8 presents a grid for top-down assessment by means of a written examination of individual translation work.

Table 8. Assessment grid for top-down assessment by means of a written examination of individual translation work

Factor	Details
Training method:	Individual translation work
Type of assignment:	Translating
Focal point of assessment:	Translation as a product, translation process (from source text to target text), translator
Assessment type:	Same as in Table 1
Assessor:	Trainer
Assessment status:	Formal
Assessment form:	Written examination
Assessment method:	Mostly analytically, possible also by positive and negative assessment
Assessment criteria:	e.g. the criteria proposed by al-Qinai (2000) and/or conformance with instructions
Assessment metric:	Number of errors/mistakes, possibly subdivided into grave, major and minor errors/mistakes
Translation theory applied:	Linguistically oriented approach
Competences assessed:	Translation competence (ISO 17100), linguistic and textual competence (EN 15038/ISO 17100), competence in research, information acquisition and processing (ISO 17100), technical competence (EN 15038/ISO 17100), cultural competence (EN 15038/ISO 17100)
Assessment requirements:	Validity: the assessment measures translating and nothing else, covers the skills necessary for translating and predicts adequate future performance accurately
	Reliability: even if there is an institution-wide set of assessment criteria and metric, supplementary measures may be necessary to guarantee reliability.
	Same as in Table 1

In this way, a grid can be composed for every assignment type, training method, etc. Grids may be combined into composite grids by factor, e.g. training method, assignment type, etc. as long as the details do not get mixed up. Grids may also be expanded or condensed depending on the situation at hand.

Feedback

Feedback in a skills lab, too, varies along a number of dimensions: form, source, object, and timing. There are various forms of feedback: oral vs. written, holistic vs. analytic, unidirectional vs. multidirectional, individual vs. group and pedagogic (aimed at improving self-regulation) vs. professional (aimed at improving service provision) (cf. Klimkowski, 2015). The feedback may come from diverse sources: clients, trainer-director of skills lab, external professional translators, external coach, management, and colleague translators/revisers. The object of feedback can be production, functioning, communication, quality delivered and learning objectives. Finally, the timing varies as well: from once, to at intervals to continuously. All these variations apply to the skills lab and could be added to the above grids. By its very nature, a skills lab offers the most opportunities for feedback.

CONCLUSION

Generally, the emphasis of quality assessment in the industry seems to be more on conformance with requirements agreed with the client, whereas in translator training this seems to be mostly on the linguistic characteristics of translations and their correspondence with those of the source text, translator training should introduce quality management and quality assurance by students into the curriculum. This can best be done in the form of a skills lab or student company. Not only may this require a curriculum adjustment, it also has its bearing on quality assessment and calls for the adoption of another focal point of assessment, notably translation as service provision as outlined in EN 15038/ISO 17100. This, in its turn, has an impact on assessment methods, criteria and metrics, which can only be different from those for translations.

Quality assessment of translations can be done by students and trainers; assessment by students is very often less reliable and objective than by trainers, although the latter may have problems with reliability and objectivity as well. This may be due to the assessment methods, criteria and metrics used, but also to other factors such as type of assignment, and translation theory adhered to. It is advisable to assess translation as a product differently than translation as a service provision because of the different competences involved.

Quality assessment is part of a system of interdependent factors. For quality assessment in a skills lab quite a few more of these factors are involved than for the translation products. For this purpose, an assessment grid was developed that can both be used for translations and other assignments. As such it may provide a basis for accreditations and allow the translator training programme to be in line with international standards for translation quality.

REFERENCES

- Al-Qinai, J. (2000). Translation quality assessment: Strategies, parameters and procedures. *Meta*, 45(3), 497-519. Retrieved July 2, 2017, from <https://www.erudit.org/fr/revues/meta/2000-v45-n3-meta162/001878ar.pdf>
- Angelelli, C. V. (2009). Using a rubric to assess translation ability: Defining the construct. In C. V. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in Translation and Interpreting Studies: A call for dialogue between research and practice* (pp. 13–47). Amsterdam: John Benjamins. doi:10.1075/ata.xiv.03ang
- ASTM International. (2006). *F2575-06 standard guide for quality assurance in translation*. West Conshohocken, PA: ASTM International.
- Catford, J. (1965). *A linguistic theory of translation*. Oxford, UK: Oxford University Press.
- CEN. (2006). *Translation services: Service requirements, EN 15038*. Brussels: CEN.
- Chesterman, A. (1997). Memes of translation. In *Translation strategies*. Philadelphia: John Benjamins. doi:10.1075/btl.22
- Colina, S. (2009). Further evidence for a functionalist approach to translation quality evaluation. *Target*, 21(1), 97–134.
- Directorate-General for Translation. European Commission. (2015). *DGT translation quality guidelines*. Retrieved June 13, 2017, from http://ec.europa.eu/translation/maltese/guidelines/documents/dgt_translation_quality_guidelines_en.pdf
- Drugan, J. (2013). *Quality in professional translation: Assessment and improvement*. London: Bloomsbury.
- Drugan, J. (2014, November). Top-down or bottom-up: What do industry approaches to translation quality mean for effective integration of standards and tools? *Translating and the Computer 36. Proceedings of The International Association for Advancement in Language Technology*.
- EMT Expert Group. (2009). *Competences for professional translators, experts in multilingual and multimedia communication*. Retrieved July 11, 2017, from https://ec.europa.eu/info/sites/info/files/emt_competences_translators_en.pdf
- Eyckmans, J., Anckaert, P., & Segers, W. (2009). The perks of norm-referenced translation evaluation. In C. V. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting Studies: A call for dialogue between research and practice* (pp. 73–93). Amsterdam: John Benjamins. doi:10.1075/ata.xiv.06eyc
- Faculty Innovation Center, University of Texas at Austin. (n.d.). *Peer assessment*. Retrieved July 6, 2017, from <https://facultyinnovate.utexas.edu/teaching/check-learning/feedback/peer-assessment>
- Gouadec, D. (2007). *Translation as a profession*. Amsterdam: John Benjamins. doi:10.1075/btl.73

- Gouadec, D. (2013). Quality in translation. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of Translation Studies* (pp. 271–275). Amsterdam: John Benjamins.
- Graham, J. F. (1985). *Difference in translation*. London: Cornell University Press.
- House, J. (2015). *Translation quality assessment. Past and present*. Oxon, UK: Routledge. Retrieved June 25, 2017, from <http://www.translationindustry.ir/Uploads/Pdf/Translation%20Quality%20Assessment.pdf>
- Hoyle, D. (2001). *ISO 9000 quality systems handbook* (4th ed.). Oxford, UK: Butterworth Heinemann. Retrieved June 11, 2017, from <http://www.pqm-online.com/assets/files/lib/books/holye2.pdf>
- Hurtado Albir, A. (Ed.). (2017). *Researching translation competence by PACTE Group*. Amsterdam: John Benjamins. doi:10.1075/btl.127
- ISO. (2005). *Quality management systems: Fundamentals and vocabulary, ISO 9000:2005(E)*. Geneva: ISO.
- ISO. (2015). *Translation services: Requirements for translation services, ISO 17100:2015*. Geneva: ISO.
- J.D. Edwards QA Form. (n.d.). Retrieved July 4, 2017, from <https://www.slideshare.net/patigalin/translation-quality-measurement2-6863666>
- Kelly, D. (2005). *A handbook for translator trainers*. Manchester, UK: St. Jerome.
- Klimkowski, K. (2015). *Towards a shared curriculum in translator and interpreter education*. Wrocław: Wydawnictwo Wyższej Szkoły Filologicznej we Wrocławiu, Komisja Nauk Filologicznych PAN, Oddział we Wrocławiu & International Communicology Institute.
- Koby, G. S., Fields, P., Hague, D., Lommel, A., & Melby, A. (2014). Defining translation quality. *Revista Tradumàtica: Technologies de la traducció*, 12, 413-420. Retrieved June 13, 2017, from <http://revistes.uab.cat/tradumatica/article/view/n12-koby-fields-hague-et-al/pdf>
- Kockaert, H., & Segers, W. (2017). Evaluation of legal translations: PIE method (Preselected Items Evaluation). *JoSTrans, The Journal of Specialised Translation*, 27, 148-163. Retrieved July 12, 2017, from <http://www.jostrans.org/>
- Koehn, P. (2017). *Statistical machine translation. 2nd Public Draft*. Retrieved November 17, 2017, from <https://arxiv.org/pdf/1709.07809.pdf>
- Larson, M. (1984/1998). *Meaning-based translation: A guide to cross-language equivalence*. Lanham, MD: University Press of America.
- Li, L., Steckelberg, A. L., & Srinivasan, S. (2008). Utilizing peer interactions to promote learning through a web-based peer assessment system. *Canadian Journal of Learning and Technology (CJLT)*, 34(2). retrieved July 16 2017 from <https://www.cjlt.ca/index.php/cjlt/article/view/26423/19605>
- LISA QA Model 3.1. (2006). Retrieved July 4, 2017, from <http://dssresources.com/news/1558.php>
- Munday, J. (2008). *Introducing Translation Studies: Theories and applications*. London: Routledge Taylor & Francis.
- Newmark, P. (1988). *A textbook of translation*. New York: Prentice Hall.

Quality and Quality Assessment in Translation

Nida, E., & Taber, C. (1969). *The theory and practice of translation*. Leiden: Brill.

Nord, C. (1991). *Text analysis in translation: Theory, methodology, and didactic application of a model for translation-oriented text analysis*. Amsterdam: Rodopi.

Nord, C. (1997). *Translating as a purposeful activity: Functionalist approaches explained*. Manchester, UK: St. Jerome.

Olvera Lobo, M. D., Robinson, B., Castro Prieto, R. M., Quero Gervilla, E., Muñoz Raya, E., Murillo Melero, M., ... Díez Lerma, J. L. (2007). A professional approach to translator training (PATT). *Meta*, 52(3), 517–528. doi:10.7202/016736ar

SAE International. (2005). *Translation quality metric SAE J2450_201608*. Retrieved July 4, 2017, from http://standards.sae.org/j2450_200508/

Segers, W., & Kockaert, H. (2016). Can subjectivity be avoided in translation evaluation? In M. Thelen, G.-W. van Egdom, D. Verbeeck, Ł. Bogucki, & B. Lewandowska-Tomaszczyk (Eds.), *Translation and meaning: New series*, 1 (pp. 69-78). Frankfurt a/Main: Peter Lang.

Thelen, M. (2005). Going European: A plea for building a network of simulated translation bureaus. In M. Forstner & H. Lee-Jahnke (Eds.), *CIUTI-Forum Paris 2005. Regards sur les aspects culturels de la communication* (pp. 219-236). Frankfurt a/Main: Peter Lang.

Thelen, M. (2016a). Quality and objectivity of testing and assessing in translator training: Dilemmas and suggestions. In M. Thelen, G.-W. van Egdom, D. Verbeeck, Ł. Bogucki, & B. Lewandowska-Tomaszczyk (Eds.), *Translation and meaning: New series*, 1 (pp. 79-98). Frankfurt a/Main: Peter Lang.

Thelen, M. (2016b). Collaborative translation in translator training. *Konin Language Studies KSJ*, 4(3), 253-269. Retrieved June 24, 2017, from <http://www.ksj.pwz.konin.edu.pl/wp-content/uploads/2017/02/KSJ-43-253-269.pdf>

Toury, G. (1995). *Descriptive Translation Studies and beyond*. Amsterdam: John Benjamins. doi:10.1075/btl.4

van der Meer, J., Görög, A., Dzeguze, D., & Koot, D. (2017). *Measuring translation quality: From translation quality evaluation to business intelligence*. De Rijp, The Netherlands: TAUS. Retrieved July 6, 2017, from <http://info.taus.net/quality-management-whitepaper?submissionGuid=3d87690e-380e-45cc-9e44-d2013331530f>

van Doorslaer, L. (2007). Risking conceptual maps. *Target*, 19(2), 217-233.

Venuti, L. (1995). *The translator's invisibility: A history of translation*. London: Routledge. doi:10.4324/9780203360064

Vermeer, H. J. (1978). Ein Rahmen für eine allgemeine Translationstheorie. *Lebende Sprachen*, 23(1), 99–102.

Vinay, J.-P., & Darbelnet, J. (1995). *Comparative stylistics of French and English: A methodology for translation* (J. Sager & M.-J. Hamel Trans.). Amsterdam: John Benjamins. (Original work published 1958)

Williams, M. (2009). Translation quality assessment. *Mutatis Mutandis*, 2(1), 3–23.

KEY TERMS AND DEFINITIONS

Feedback: Comments given/received on a product or service delivered.

Focal Points of Assessment: Perspectives for judging and measuring quality.

Quality Management: Measures taken in an organization to stimulate, regulate, and keep constant the quality of products and services delivered.

Translation Quality: The degree in which a product or service delivered conforms with requirements because of its inherent characteristics.

Translation Quality Assessment: The process of checking the quality of translation(-related) products or services delivered.

Translation Quality Assurance: Measures taken in an organization to guarantee translation quality.

ENDNOTES

¹ “Inherent” is opposed to “assigned”, and ... means “existing in something, especially as a permanent characteristic” (ISO 9000:2005(E), p. 7). This chapter will not pay any attention to this contrast, since it does not affect the essence of the definition of quality.

² The European standard EN 15038 does not define the concept of quality.

³ See Munday (2008) for another classification.

⁴ Translation Service Provider.

⁵ A skills lab is an in-house simulated translation bureau, not to be confused with placement. A skills lab takes place inside the training institute, whereas placement takes place outside the institute. In this chapter, the skills lab at the Maastricht School of Translation and Interpreting of Zuyd University of Applied Sciences in The Netherlands is taken as an illustration of the introduction of quality management and quality assurance by students into the curriculum and forms the input for the proposed assessment system and grids. In Maastricht, the skills lab is a regular, intra-curricular part of the professionally oriented curriculum and is mandatory for all students. The Maastricht skills lab is a simulation of a real translation bureau, but is run and staffed by fourth-year students under the supervision of a trainer-director and with the input of a number of professional translation bureaus and the guidance of an external professional coach (for a description see Thelen, 2005). A skills lab is typical for professionally oriented translator training institutes, but is by no means the only option to implement professionalisation (Kelly, 2005; Gouadec, 2007; Olvera Lobo et al., 2009).

⁶ A student company in this sense is a company that is part of the curriculum and earns credit points, not to be confused with a student company that is extra-curricular and does not earn credit points.

⁷ These institutes are members of INSTB, the International Network of Simulated Translation Bureaus (cf. www.instb.eu).

⁸ Both in the translation industry and translator training, quality assessment is also called quality evaluation. Kelly (2005) reserves the term evaluation for the interrogation of training institutes in the context of accreditations on the training offered and the results achieved. This chapter will use the term quality assessment (allowing for evaluation in direct citations).

⁹ Collaborative translation is further discussed in Thelen (2016b).

Quality and Quality Assessment in Translation

- ¹⁰ LISA (the Localization Industry Standards Organization) ceased to exist in February 2011.
- ¹¹ “Peer assessment grades are recommendations only, and the instructor makes final grading decisions” (Faculty Innovation Center of the University of Texas at Austin, n.d.).
- ¹² A combination of EN 15038/ISO 17100 and EMT competences will serve as guideline here.
- ¹³ Table 5 does not identify criteria for assigning the scores 1 to 5 to translations produced in the assessment metric field. These will depend on the particular translation theory applied and the overall assessment system of the training institute in question. The range from 1 to 5 is flexible and is set locally depending on the organisation of the skills lab (the number of trainer-directors and the number of translations to be assessed).

Chapter 2

How to Put the Translation Test to the Test? On Preselected Items Evaluation and Perturbation

Gys-Walt Van Egdom

Zuyd University of Applied Sciences, The Netherlands

Heidi Verplaetse

KU Leuven, Belgium

Iris Schrijver

University of Antwerp, Belgium

Hendrik J. Kockaert

KU Leuven, Belgium

Winibert Segers

KU Leuven, Belgium

Jasper Pauwels

University of Antwerp, The Netherlands

Bert Wylin

KU Leuven, Belgium

Henri Bloemen

KU Leuven, Belgium

ABSTRACT

Reliable and valid evaluation of translation quality is one of the fundamental thrusts in present-day applied translation studies. In this chapter, a thumbnail sketch is provided of the developments, in and outside of translation studies, that have contributed to the ubiquity of quality in translation discourse. This sketch reveals that we will probably never stand poised to reliably and validly measure the quality of translation in all its complexity and its ramifications. Therefore, the authors have only sought to address the issue of product quality evaluation. After an introduction of evaluation methods, the authors present

DOI: 10.4018/978-1-5225-5225-3.ch002

How to Put the Translation Test to the Test?

the preselected items evaluation method (PIE method) as a perturbative testing technique developed to evaluate the quality of the target text (TT). This presentation is flanked by a case study that has been carried out at the University of Antwerp, KU Leuven, and Zuyd University of Applied Sciences. The case study shows that, on account of its perturbative qualities, PIE allows for more reliable and more valid measurement of product quality.

INTRODUCTION

Translation assessment and evaluation are two subdomains of Translation Studies that have been proven to be fraught with obstacles. In the past, the obstacles seemed mainly problematic to translation service providers (i.e. translation agencies and freelance translators), whose primary objective it is to sell (high) quality products and services as a means to attract new clients and maintain customer loyalty. The discipline of Translation Studies has kept quality at a safe distance for a number of years.

Before turning the attention to the sudden revival of academic interest in quality, it is worth pausing to consider the underlying reasons scholars had for skirting the issue. Quality was the main reason why the discipline had started off on the wrong foot. In the 1950s and 1960s, the heyday of what some would call “Linguistic TS”, quality had been placed at the top of the agenda. The attempts of the early scholars to pinpoint the characteristics of an equivalent translation resulted in highly subjective and prescriptive ideas about translation quality. Still, this is not all there is to it: the prescriptivist tendencies even led to the unjustified equation of evaluative judgments and categorial judgments; in other words: texts that did not conform to a scholar’s theoretical preconceptions of a quality translation were simply denied the full status of “translations” (cf. Even-Zohar, 1990; Chesterman, 1997, p. 63). Although Popovič has been hailed by many scholars (amongst others by Holmes, 1988) as a renewer of TS, his lemma “Translation”, in his critically-acclaimed *Dictionary for the Analysis of Literary Translation* (1976) – originally published as an addendum in *Téoria umeleckého prekladu* (1975) and translated as such in *La Scienza della Traduzione* (2006) – bears witness to this unjustified equation of translation and high-quality translation. As an advocate of stylistic equivalence, he defines translation as the “passaggio dell’invariante da un testo all’altro osservando in modo ottimale le peculiarità stilistiche e specifiche del prototesto [etc.]” (2006, p. 174).¹

In the last quarter of the twentieth century, TS scholars have done their utmost to redress this unjust situation and to raise the profile of TS. First and foremost, this meant that emphasis needed to be placed on empirical findings. It soon dawned upon the new generation of scholars that the quality of a translation is always relative to the social-cultural context in which a translation is embedded. At the end of the century, no or hardly any doubt was entertained about translation quality: a correct translation was “the one that fits the correctness notions prevailing in a particular situation, i.e., that adopts the solutions regarded as correct for a given communicative situation, as a result of which it is accepted as correct” (Hermans, 1991, p. 166). It was up to TS to provide a full and unbiased overview of the norms prevailing in cultural systems of all times, all around the world.

Around the turn of the millennium, the call for quality became ubiquitous both in the field of translation and in that of TS. Translation service providers thus found a somewhat unexpected ally in their quest for quality: scholars. It had become clear that, although it need not work to the advantage of the theoretical branch(es) of TS, these scholars still had a vested interest in the assessment and evaluation of translation for the simple reason that most scholars are also teachers.

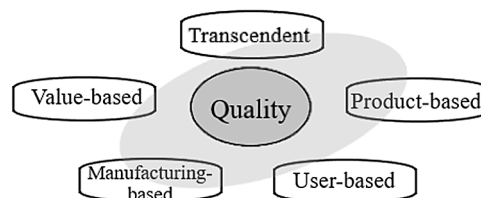
In recent decades, translator training has made steady progress on the road to what is called “professionalisation”. The early functionalists had seized upon the idea of translation as communicative action. It would not take them long before they started looking for effective ways to translate their ideas into teaching and learning strategies and to mesh them with plans for curricular redesigning (Hönig & Küßmaul, 1982; Nord, 1988). In the 1980s and 1990s, Delisle greatly contributed to the professionalisation movement: he repeatedly placed emphasis on the need for “objectifs d’apprentissage” (1980, 1990, 1993, 1998) in translator education. With his “learning outcomes”, Delisle presaged the competence discussions that would soon change the face of translator training (Schäffner & Adab, 2000; PACTE, 2002, 2003, 2016a, 2016b; Alves & Hurtado-Albir, 2017; Hurtado-Albir, 2007, 2017). In order to become a successful translator, a trainee must acquire the competences relevant to translation. This implies that there must also be means to assess and evaluate the acquisition of these competences and even to gauge competence levels.

With the competence approach in full swing, the bar has been raised even higher for translator education. First of all, the European Master’s in Translation (EMT) network, officially launched in 2009, has created a gold standard for translator training. Clearly going beyond the traditional ken of translator training, the network has set high priority on translation services (EMT Expert group 2009; cf. Klimkowski, 2016, 2017). The Directorate-General for Translation (DGT) of the European Commission has noted that there is still a “skills gap” separating training and practice (cf. Thelen, 2016a, 2016b). If this gap is to be narrowed or, ideally, closed, due consideration should be given to aspects of service provision (cf. ISO, 2015). Hence, it is imperative for translation trainers to attain firm ground for understanding and, ultimately, assessing and evaluating those service aspects. Which criteria should be applied in this context? One would be inclined to say that the translation industry holds the answer. However, due to ongoing diversification and, not unimportantly, computerisation, the industry seems to grapple more than ever with the issue of quality (cf. Drugan, 2013; Saldanha & O’Brien, 2014; Fields, Hague, Koby, Lommel, & Melby, 2014; Koby, Fields, Hague, Lommel, & Melby, 2014; Melby, Fields, Hague, Koby, & Lommel, 2014).

To give the reader a rough idea of the complexity of defining translation quality and of the scale of the research needed to develop (reliable and valid) tools to assess and evaluate translation quality in all its complexity, Garvin’s five general approaches to quality are reproduced below (Figure 1; Garvin, 1984, 1988; Tenner & DeToro, 1992; Koby et al., 2014). These approaches will be briefly related to translation.

Translators and translator trainers holding a transcendent view of translation quality seem unable to define quality itself or, even, to sum up the concrete characteristics of a quality translation, yet they immediately recognise a translation as being of good or excellent quality. Those holding a product-based view strive for the quantification of quality: by identifying a number of characteristics or attributes, one will be able to tell whether a translation passes muster. When the quality of a translation hinges entirely on the satisfaction of end-users’ preferences, the view is called “user-based”. In the translation industry,

Figure 1. Garvin’s five approaches to quality (1984)



How to Put the Translation Test to the Test?

the perceived quality of translation (services) has received considerable attention in the past few years (cf. Massardo, Van der Meer, & Khalilov, 2016, pp. 84-86). The conformity of translation to service requirements plays a crucial part in manufacturing-based approaches to quality. The above-mentioned ISO standard (ISO, 2015) sets out the requirements for translation services. Disputes between translation service providers and their customers are frequently settled on the basis of these requirements. As the name suggests, value-based approaches zero in on quality-price ratios. The one question that needs to be answered in this context, is: does the translation offer value for money? An affirmative answer suggests that the translation is up to scratch. When observed through the prism of quality, one soon understands why, despite the countless obstacles one is bound to incur, the themes of assessment and evaluation have been gaining tremendous momentum in recent years. Whereas until (fairly) recently translation trainers tended to rely predominantly on what Eyckmans, Anckaert and Segers (2009) have come to call “intuitive-impressionistic” methods (p.73) and focused mainly (if not solely) on the summative assessment and evaluation of products (e.g., student translation), their focus has now gradually shifted toward analytical, formative and, sometimes even, inclusive means of assessment, which serve a diagnostic purpose and should allow students and teachers alike to monitor progress along learning curves.

As a result, an unparalleled number of methods, models and metric systems have been proposed and a few have even marketed in the 21st century (cf. *infra*). With the proliferation of methods, models and metrics, all having their affordances and limitations, the need for empirical evidence for the quality of translation tests has been professed in TS (cf. Waddington, 2001, 2004; Martínez Melis & Hurtado Albir, 2001; Anckaert & Eyckmans, 2006; Van de Poel & Segers, 2007; Anckaert, Eyckmans, & Segers, 2008; Angelelli & Jacobson, 2009; Eyckmans et al., 2009; Williams, 2009; Kockaert & Segers, 2012; Koby & Melby, 2013; Mariana, Cox, & Melby, 2015; Eyckmans, Anckaert, & Segers, 2016). When dealing with testing quality, a few relevant questions immediately spring to mind: Is a translation test reliable? Is it valid?; How can one even attest to the reliability of a translation test? and How can one attest to its validity? Time has taught that there is no easy answer to all of these questions. What is self-evident to all translator trainers, even to those with the barest experience in assessment and evaluation, is that it is extremely difficult to apply standard testing taxonomy, such as Bloom’s widely-used taxonomy of educational objectives (1972), to phenomena like translating and translation. All the more so because concerns about professionalisation and employability are prompting translator trainers to seek ways to bridge the gap between training and practice. As was mentioned above, translator trainers would do well to always take heed of not only educational objectives, but also of professional requirements (cf. Thelen 2016a, 2016b).

In light of the above, it is not surprising that, in spite of the firm resolution to include more aspects of assessment and evaluation in research projects in the near future, the authors of this chapter have been urged to narrow the scope of their study, and confine themselves to the quality of translation testing (rather than to the testing of translation quality). More concretely, the authors hope to forge a persuasive argument for the use of the preselected items evaluation method (or PIE method) in translation assessment and evaluation. It is their contention that the PIE method offers bright prospects for (more) reliable and (more) valid testing. In order to properly understand and evaluate the potential of this analytical testing method, it might prove useful to cast a brief glance at the way the PIE method is embedded in the context of translation assessment and evaluation. In the next section, a brief overview will therefore be given of prominent product quality testing methods. In the following section, PIE will be brought to the fore: its major strength will be highlighted by dint of comparison. PIE tests are presented as perturbative theories. In order for the reader to glean an idea of how a perturbative approach might contribute to the

establishment of reliability and validity in translation testing, a case study, conducted by researchers from three different institutions (University of Antwerp, KU Leuven and Zuyd University of Applied Sciences) will be presented.²

At the end of this chapter, it will become clear why the authors maintain that PIE provides an impetus to testing quality on account of its so-called perturbative traits.

EVALUATION

It is tempting yet most undesirable to employ the terms “assessment” and “evaluation” interchangeably in a didactic setting. It would be a truism to claim that the terms are incontestably related, and that, in the vast majority of cases, they can even be seen as concomitant terms – virtually every assessor passes an evaluative judgment on a TT and virtually every translation evaluation rests on an assessment made by the evaluator. However, this chapter will not see assessing and evaluating as synonyms and make a careful distinction between them.

A possible definition of “assessment” would be: the process of measuring and understanding the effectiveness of teaching and learning strategies and methods. Good assessment methods provide information not only on the effectiveness, but also on the improvement of learning and teaching. In translator training, for instance, it is crucial to have a good handle on students’ competences. By assessing behavioural indicators of competence acquisition – indicators that are ideally linked-up with competence descriptors – one can get an idea of the students’ progress. By assessing a group of students, one can also tell whether teaching methods are geared to the needs of students.

Evaluation, on the other hand, can be defined as a process involving a value judgment on a performed task. In the case of translation, this task is the product, the process or the service (Kockaert & Segers, 2017). A value judgment can always be placed on a scale, marking the difference between optimal, passable and poor performances. In the vast majority of cases, the appreciation of the evaluator takes on the form of a percentage (e.g. 60%), a grade (e.g. 3/10) or an evaluative letter (e.g. A) awarded to the student(s). Evaluation serves the purpose of marking and justifying the difference between passing and failing.

In pedagogy, assessment is seen as a formative activity and evaluation as a summative activity. As the name “PIE” suggests, the authors’ focus will fall on translation evaluation. Still, it should be noted that the method can easily be turned into a didactic method, making it a tool for teaching and assessment as well. In point of fact, the authors’ experience has taught that student interest in evaluation is piqued through the “didactisation” of the PIE method.

EVALUATION METHODS

Before attention is drawn to the evaluation methods presently at hand in translator training, it is worth casting a quick backward glance at the way both evaluation and assessment have been embedded in TS. As is well-known, Holmes prepared the ground for TS in a paper that was first delivered in 1972. In this paper, entitled “The Name and Nature of Translation Studies” (Holmes, 1988), he proposes a three-pronged scientific approach to translation. Translation assessment and evaluation are part and parcel of the “applied” approach to translation. Still, in his paper, Holmes makes no mention at all of translation

How to Put the Translation Test to the Test?

assessment; he only suggests that the efficacy of translation in language learning should be “assessed” (p. 78). Evaluation receives more of Holmes’ attention, yet it is placed neatly under the heading of “translation criticism” (p. 78), which is set apart from “translator training” (p.77). In other words, research on “evaluation” was meant to cover evaluation of professional translation(-related) practices. Does this mean that, according to Holmes, there is no room for assessment and evaluation in translator training? Probably not: in his section on translator training, Holmes seems to have included “testing techniques” to buoy up assessment and evaluation in training (p. 77).³

Despite the fact that Holmes only paid scant attention to assessment and evaluation in training, his passage on evaluation serves as an excellent point of departure in the discussion of evaluation methods. Holmes – who, by the way, did not believe it was the task of TS to aid and abet translation professionals in their struggle for objective quality standards – contends that “the activities of translation interpretation and evaluation will always elude the grasp of objective analysis to some extent, and so continue to reflect the intuitive, impressionist attitudes and stances of the critic” (p. 78). Almost half a century later, the idea that it is impossible (or, at most, very difficult) to shake off subjective influences in assessment and evaluation still rings true. What is different today is the attitude and stance toward evaluation and assessment: even those who are inclined to highly subjective evaluation are now more sensitized to the fact that their evaluation is subjective.

The most important change that has been witnessed in translation evaluation in the past decades has been brought about through the introduction of analytical models. It is no coincidence that, in the above-cited passage, Holmes makes reference to intuition and impressionism, two terms the reader has stumbled upon earlier in this chapter. Throughout the 20th century, holistic evaluation approaches held sway in translator training as well as in the translation industry (cf. Beeby, 2000, p. 185; Valero-Garcés, Vigier, & Phelan, 2014). The reason why these methods are called intuitive and impressionistic, is because the assessment of the quality of a translation is based on an overall impression of the translation – the TT taken as a whole (Mariana et al., 2015, p. 155; Eyckmans et al., 2016, p. 75; Kockaert & Segers, 2017, p. 149). Although disagreement amongst researchers remains and further investigation is warranted, some studies have suggested that there is or, at least, can be a high correlation between scoring in holistic and in “reasoned” evaluation (Beeby & Martínez, 1998; Waddington 2001, pp. 315-317). As Beeby (2000) rightly points out, one should take into account the “experience” of the evaluator (p. 85; cf. Beeby & Martínez 1998); a seasoned evaluator is trained to distinguish between good and bad translations and is more able to pass a seemingly intuitive judgment on a TT (cf. supra: “transcendent view”). As a consequence of the above-mentioned tenets and outcomes, there are still scholars championing this approach or trying to salvage it by proposing a combined evaluation (e.g. Waddington 2001; 2004, p. 34 Garant 2009). This being said, it is self-evident that the concept of holistic evaluation itself is difficult to sell. Although a method should not be disparaged all too flippantly for being a weak “defence mechanism” against student criticism, every holistic evaluation is founded on little more than an appeal to authority (cf. Anckaert & Eyckmans 2007, pp. 54-55). In other words, the holistic method does not provide unequivocal justification of decisions made in the evaluation process (cf. Waddington, 2001).

As indicated, analytical evaluation seems to have outflanked holistic evaluation in the past decades (cf. Saldanha & O’Brien, 2014, p. 101). The first thing to note apropos analytical evaluation is that it is an umbrella term under which a variety of methods, models and metrics can be grouped. Although the majority of analytical methods are more time-consuming than holistic methods, analytical evaluation has the advantage that evaluators and (student) translators alike stand poised to gain a more tangible sense of what is right and/or wrong in a translation (Kockaert & Segers, 2017, p. 150). The following defini-

tion of Mariana et al. (2015) sheds light on the nature of analytical evaluation: “[the analytical method] is a way to assess the quality of a translation by looking at segments of the text [...] and awarding or deducting points to the overall score of the translation based on whether each text unit meets certain criteria” (p. 155).

A landmark analytical model is the LISA QA model (Localization Industry Standards Association 2004).⁴ In this customizable model, which is still in common usage in the translation industry, the quality metrics of various multinational companies (IBM, Rank Xerox, Microsoft...) were brought together so as to provide a solid basis for more objective and, what is more important, more justifiable evaluation of translations (cf. O’Brien, 2012). The QA model would catch on quickly and was even put into place in translator training, where there was growing awareness of the necessity of attuning to the needs of the profession. Despite its successes, the model has been increasingly superseded by other customizable models and metrics, such as the Dynamic Quality Framework (DQF) and the Multidimensional Quality Metrics (MQM). However, the imprint of the LISA model has proven to be indelible: the error categories advanced in the DQF (O’Brien, 2012; Görög, 2014) and MQM (Lommel, 2014; Lommel, Burchardt, & Uszkoreit, 2014; Mariana et al., 2015) bear strong resemblance to the LISA error categories (e.g. accuracy, style, and terminology).⁵ Even though these days the analytical method may well be considered the optimum course for action in translation evaluation, it is still widely recognised that this method is not without flaws. Wishing to abide by the basic principles of test designing, principles that do not allow for the simple registration of errors regardless of the course content that is being tested or of the intended learning outcomes, some scholars have tried to remedy the demerits of analytical testing by proposing a rubricated approach. According to Colina (2009), the rubricated approach is “componential because it evaluates components of quality separately” and, “functionalist because evaluation is carried out relative to the function and the characteristics of the audience specified for the translated text” (p. 240). The componential aspect is highlighted in the form of (error) categories, called rubrics. These rubrics often call to mind the typology in analytical evaluation. A fitting example of a rubric method that lays emphasis on known error typologies is Wang’s method (2015). Her tests first yield separate scores on 1) accuracy, 2) register and terminology, 3) grammar, cohesion, coherence and organisation of work, and 4) spelling, accentuation, punctuation, including formatting for local conventions and regional standards. The awarded scores are later aggregated to obtain a total score. How the “rubric method” is to be conceived of remains open to discussion: to scholars like Colina (2009), it represents a viable “third way”, others just see it as a dynamic method that is to be placed on a continuum between holistic and analytical evaluation (amongst others Mariana et al., 2015).

As is evidenced by the volume of published research on evaluation, of which the contributions in this chapter and even in this book will recap and represent only a small sample, some serious headway has been made in recent years. Still, scientifically sound answers to the question as to how translation evaluation can be freed from subjective influences and how this can be achieved without evaluation becoming a fully automated process (in which the evaluator has no say at all) seem to remain out of reach. The authors contend that these questions are now more vexing than ever, because, all things considered, they remain the only key to assuring the quality of a test design, i.e., the key to reliable and valid testing.

Although the noble and, above all, painstaking attempts of TS scholars does seem to have led to a more critical and systematic approach to evaluation (and assessment), one cannot but conclude that, notwithstanding the successes holistic and analytical methods have met with in, *inter alia*, Waddington’s research (2001), the scientific evidence still bodes ill for the vast majority of evaluation methods, models and metrics, as they fail to provide the much-needed fillip to guarantee (more) reliable and (more) valid

How to Put the Translation Test to the Test?

testing. A case in point is the analytical grid that has become common currency in professional practice and translator training. In this context, the following passage is well worth quoting at length:

while a list of specific errors, penalties, weightings, etc. has the allure of being somewhat objective, there are a number of problems associated with error typologies. They can take a lot of time to apply [...]. Even though error categories can be well-defined, it is often a matter of subjective judgement as to whether an error falls into one category or another [...] Likewise, the classification of an error as major or minor is subjective, even if definitions are provided (Saldanha & O'Brien, 2014, pp. 101-102).⁶Over the past decade, the authors have been working feverishly to overcome the setbacks and shortcomings detailed above. In the next paragraph, the PIE method will be presented as a quite successful attempt to neutralize the inevitable subjective influences in translation evaluation. It is not superfluous to spell out that the PIE method is by no means presented as a "third way", nor is it devised as a "middle way" between holistic and analytical evaluation. The PIE method is considered to be an analytical model. Furthermore, PIE can be made compatible (if need be) with analytical assessment grids (e.g., DQF, MQM).

PIE METHOD: BACKGROUND

At first blush, the PIE method appears to be in broad agreement, to say the least, with the method that has been dubbed the "Rich points method". The latter method was ushered in by functionalist scholars, who tended to take a firm stance against equivalence-based approaches to translation. One of the theorists advocating the use of this method to evaluate translations was Christiane Nord. In her seminal work *Textanalyse und Übersetzen* (1988), Nord endorsed the functionalist project fully, yet she refused to do away with the source text. In point of fact, the source text and its socio-cultural function served as a point of departure in her model for translation-oriented text analysis. According to Nord, a translation-oriented text analysis yields invaluable information about and insight into the prospective translation processes and, more precisely, the obstacles that are likely to stand between the translator and a fully functional TT.⁷

Whenever there seems to be a lack of fit between the communicative goal of the TT and the means at the disposal of the translator, the likelihood that translation errors will occur increases dramatically (Nord, 1987, 1988). For that reason, Nord argues, it would be highly rewarding for evaluators to home in on those problem areas she has come to call "rich points" (Nord, 1994, 1997, 2002, 2011a). Nord's ideas laid the foundation of an evaluation method that isolates elements in the source text (e.g., words, word groups, sentences) that are likely to present difficulties to the (aspiring) translator and that only pays heed to the translation of those fragments in the evaluation process.

In the first years of its application, the rich points method was primarily valued as a means of facing the lack of professionalism in translator training squarely: the method placed strong emphasis on the translation brief, a brief that was to mirror authentic commissions/assignments, and, ultimately, pursues the goal of valid testing. In the 1990s and early 2000s, the method would follow the fashion of educational sciences. This meant that the evaluation method was championed, again by Nord (2011b), because it enabled teachers to relate test items to desired learning outcomes (*Lernziele*). Around that time, the method was picked up by the PACTE group, who have entertained the wish to even relate items to translation competences (2008, 2009, 2011; cf. Castillo Rincón, 2010, 2015; Hurtado Albir, 2017, p. 109, pp. 357-370).

Years of experience and hindsight have taught that there is also another reason why the rich points method has steadily gained popularity, even beyond the field of translator education. The rich points method has an edge on other analytical evaluation methods because, like holistic evaluation methods, it is (potentially) time-saving.⁸ Although it seems not uncommon to flank the evaluation of rich points with more traditional holistic methods, a great deal of time can be won by applying the rich points method alone. This is certainly the case when one is burdened with the task of correcting large quantities of versions of one and the same source text.

Still, there appears to be no reason to put the rich points method on a pedestal. As reliability and validity do not necessarily increase considerably when the method is applied, the method can easily be subjected to stern criticism. The rich points method relies too heavily on the expertise of the evaluator. The most important stage at which the evaluator's influence is felt, is the preparation phase, where the evaluator is charged with the responsibility of selecting the rich points in the source text. Who is to tell whether the selected items, no matter how carefully they have been picked out, are indeed representative of the difficulty of the translation task? Furthermore, the rich points method itself leaves as much room for error interpretation and individual weighting as most other analytical methods. It is up to the evaluator to determine which translation variants are correct and which are not; and, what is worse, it is up to him/her to judge the severity of errors made. Put to the test, a rich points translation test is still flawed in some respects.

This being as it may, the authors have recently come to the conclusion that the rich points method can stand evaluators and (aspiring) translators in good stead after all. This conclusion was reached in a roundabout way. When one traces the origins of the method under review, one soon stumbles upon a method bearing the somewhat obscure name "Calibration of Dichotomous Items". The CDI method was thought out by Eyckmans et al. (2009). In an attempt to rid translation evaluation of subjective influences, they developed a method that relies on *est material*, i.e., multiple translations of a single source text. The method itself is much akin to methods that are presently applied in machine translation engineering, as it follows some predictive procedures that are common in natural language processing ("deep learning"). In essence, it is aimed at identifying deviations in a set of ("pre-test") translations. These deviations are indicative of the slip-ups in translations that are likely to have discriminating power. This implies that, in contradistinction to the rich points method, multiple translations are first taken into account and the ensuing item selection is (most probably) a representative sample of mistakes made in the translation. It is easy to see why the CDI method would be kindly looked upon: it constitutes a major contribution to more objective translation testing, since it excludes the evaluator from the item selection process. Still, the CDI method has its demerits. Although the CDI method yields invaluable information about the qualitative differences between translations, and possibly between translators, it does not appear to mesh well with the widely-entertained aspiration to develop tools to measure learning outcomes and translation competence acquisition. The assessment and evaluation of learning outcomes and competences seem to call for criterion-based testing methods; CDI is a norm-reference method. This would suggest that, despite the increased testing reliability, the construct validity of the CDI method hangs in the balance. Furthermore, the single most important decision, the decision on the correctness of translation solutions, still lies with the evaluator.

It stands to reason that the CDI method offers a highly interesting vantage point from which translation quality can be scrutinized. Still, for the reasons summed-up above and other reasons, the authors have grown increasingly uncomfortable with its use, especially in the context of translator training. Out of

How to Put the Translation Test to the Test?

the CDI method evolved a method that would be coined the PIE method (Kockaert & Segers, 2014). At the face of it, the method runs counter to the prevailing opinion of those who seek more objective ways to evaluate translation. Having realized that the evaluator cannot possibly be taken out of the evaluation process without fully jeopardizing the validity of a test, the authors have decided to make hay of his/her somewhat undesired presence. This is where the rich points method was brought into play. By combining a classical rich point approach with the calculations made in the CDI process, the developers of the PIE method believe to have brought together the major strengths of criterion-based and norm-referenced evaluation. How so? Owing to the influences of rich point-based approaches, the PIE method is the formal solution that is most propitious to the problematic issues of finding ways to make evaluation less time-consuming and interlinking testing and learning outcomes. Owing to the influences of the CDI method, the PIE method allows for a deeper and, above all, computationally tractable understanding of translation evaluation. What is more, an increasingly convincing body of evidence (Kockaert & Segers, 2012, 2014, 2017; Segers, 2012; Segers & Kockaert 2016) suggests that the PIE method contributes greatly to intra-rater and inter-rater reliability.

This is all the more true, since the method has spurred the interest of the software engineers of Televic Education to develop translationQ. The evaluation module of the tool (evaluationQ) allows evaluators to hold a tight rein on the correctness of items, thus fully guaranteeing intra-rater and inter-rater reliability, and to speed up the evaluation process significantly. Furthermore, translationQ is believed to benefit the translator training process in other ways (speeding up revision processes, keeping tabs on error frequency, etc.). Far from qualifying the development of the evaluation module as a simple means to valorise research, the authors see the tool as a prerequisite for making steady progress in furthering the understanding of the methodological implications of translation testing in general and of PIE testing in particular, as well as in finetuning the PIE method itself, so as to make it as reliable and as valid as possible.

PIE METHOD: A PERTURBATIVE APPROACH

The question that begs answering now is: how does it work? The answer depends on the application of the method. Recently, the term “PIE Light” was coined by Kockaert and Segers (2017, p. 153). This variant is a potential source of misunderstanding and confusion and will fall outside the scope of this article.⁹ In this contribution, a fully-fledged PIE test will be showcased with the explicit aim of interlinking PIE and perturbation, a term that has gained currency mainly in the exact sciences (e.g., pure mathematics, quantum physics, and quantum mechanics). The term “perturbation”, as employed in this chapter, comprises the techniques and methods used to approximate a solution to a problem to which the real or full solution cannot be offered “in closed form” and, therefore, appears to be forever beyond reach. This linkage can be particularly insightful in the specific context of item selection. How is the rough idea underlying perturbation and perturbative theory to be understood? And how can the linkage inhibit a fruitful discussion in translation evaluation?

In perturbative theories, like the ones advanced in quantum mechanics, a complex problem is broken down into solvable and perturbative components. In order to obtain an approximate solution to the problem in question, the solvable components, always surefire solutions to (somewhat) similar problems, are first brought into play. In other words, an exact solution – if “exact” is the right word – is taken as a point of departure. Pushing forward to the approximate solution, “small terms” are later added to or subtracted from the exact solution (cf. Murdock, 2014).

An excellent example of a perturbative solution is the equation of the earth's motion through the solar system. One should recall that the motion of all objects is dependent of gravitational influences. What makes the equation so complex, is that all things exerting a gravitational pull are to be added to the equation. Therefore, it is of little surprise that the equations found in physics are beyond the grasp of even the most seasoned theoretical physicists. At this point, Greene explains in his book on string theory, perturbation comes in handy:

[W]e can [still] predict the motion of the earth through the solar system with great accuracy by making use of a perturbative approach. The enormous mass of the sun, in comparison to that of every other member of our solar system, and its proximity to the earth, in comparison to that of every other star, makes it by far the dominant influence on the earth's motion. And so, we can get a[n] estimate by considering only the sun's gravitational influence. [...] We can [also] refine this approximation by sequentially including the gravitational effects of the next-most-relevant-bodies (2011, pp. 289-290).

In this case, the gravitational influence of the sun represents an exact solution, and the “effects of the next-most-relevant-bodies” are the small terms that are introduced to get an approximate idea of the motion of the earth through our solar system.

The authors argue that there are striking similarities between the perturbative approach described above and PIE testing. In PIE testing, the quality of the translations of a given source text poses a problem. To solve this problem, the evaluator(s) proceed(s) along the lines of the rich points method. In other words, the moment the source text is selected, a search for appropriate items is carried out by the evaluator/s (ideally plural). It should be pointed out that, in PIE testing, item selection need not be restricted to “word” or “word group” selection. In PIE testing, the appropriateness of selected items hinges for the most part on the translation brief, i.e., the specific translation requirements. PIE is, thus, primarily a criterion-based analytical approach. When operating in the context of translator education, the brief should be in agreement with the desired learning outcomes and, if at all possible, with the specific translation competences that require testing. A PIE test can make minimal claims to content validity when these conditions are fully met.

Once the selection has been made, the good and passable translations of the selected items are summed up, possibly even looked up (in past tests, translation corpora or error memories (cf. Bowker, 2001)) and a list of satisfactory solutions is finally drawn up. Taken together, the preselected items and the (good and passable) translation solutions, which are basically the result of accumulated experience and are, ideally, propped up by empirical data, constitute what has been called an “exact solution”. This solution will probably yield an estimate of the quality of the translations under review, albeit a rough one.

At this point, the test can only be qualified as a “tentative theory”.¹⁰ The items and respective solutions may never be seen as a single touchstone for testing quality. This means that, as an evaluator, one is obliged to put the test itself to the test. Furthermore, one should be willing to refine or even reject the test design in accordance with the results yielded by this test. To know how one's tentative theory fares in a specific test setting, PIE requires that a critical eye is cast over the translation solutions of the items in the test translations. In this phase, the items are corrected, as in standard test situations: all translation solutions corresponding to the answers that have been formally drawn up by the evaluator(s) are marked as correct. But this is not all: after close inspection of the (allegedly) “incorrect” answers, the list of good and passable solutions is usually extended. By the extension of the list and, thus, the refinement of the test, awareness is shown of the evaluator(s) own limitations: although an extensive list of good and

How to Put the Translation Test to the Test?

passable solutions may be produced, heed is always to be taken of the strokes of genius on the part of (aspiring) translators. On the basis of the extended and refined list of solutions, the correction is carried out one more time. When performed meticulously and correctly, this triple correction mechanism gives a major impetus to reliability in translation testing, or, to be precise, to intra-rater and inter-rater reliability. In other words, PIE testing adds to the objectivity, consistency, reproducibility, stability and transparency of translation testing. PIE guarantees the equal treatment of candidates, by evaluating each of them on exactly the same items, and by helping the evaluator(s) keep a tight rein on the correctness of solutions, allowing for an accurate distinction between correct and incorrect translation solutions.

Despite its increased reliability, the translation test does not yet meet the threshold of quality at this stage. Once the tests are corrected and the list of solutions is completed, the test is still little more than a tentative theory. One still cannot tell whether the selected items are up to standard, i.e., whether they are sufficiently difficult and have enough discriminating power to allow for a sound evaluation. In other words: PIE cannot be completely set or told apart from rich point testing. To get an idea of the “minimum number of items needed for a desired level of score reliability”, it is recommended that the p value and d indices be calculated for each item (cf. Lei & Wu, 2007, p. 527). The docimological p value, which is to be distinguished from the p-value that is often calculated in scientific research to determine significance, is basically the percentage of candidates that formulated a correct answer to the test item and is, as a logical consequence, indicative of item difficulty. The d index, which is indicative of the discriminatory power of items, is based on bottom group and top group averages (cf. Matlock-Hetzel, 1997). Good items generate a p value that lies somewhere between 0.27 and 0.79 and a d index no lower than 0.29 (Ebel, 1954; Wiersma & Jurs, 1990). Items with divergent values are in most cases eliminated. These docimological data allow the evaluator to measure the quality of the test (design).

In many cases, the approximate solution is found after the calculation of p values and d indices. However, sometimes too many items should be eliminated by reason of divergent values. In these cases, one is left with too few items, too little critical mass, to assure the quality of a PIE test. The authors hasten to point out that this does not mean that the PIE test itself is flawed. The p values and d indices are norm-referenced scores. Divergent values might suggest that the group composition leaves much to be desired. Whatever the cause, when too many items need to be eliminated on grounds of docimologically recurrent observations, this is a strong indication that the test is still lacking in refinement. This means that one’s understanding of the (quality of the) test is still too coarse to provide solid grounds on which to pass a sound evaluative judgment. This need not necessarily pose a problem: some will even judge it best to leave the selected items or the items with a satisfying p value in place. However, when the stakes are high – for instance, when a decision has to be made on whether credits are to be awarded or not – one is urged to address the issue by seeking new items. A manual search for new docimologically justified items can but need not be time-consuming, as relevant items might have come to light in the correction phase. New technologies, like leading edge parsers, however, offer solutions that might make light or lighter work of item replacement. By replacing the bad items with sufficient new items, an approximate solution can be found, i.e. test quality is assured.¹¹

CASE

A case study is provided in this section to illustrate how PIE is believed to assure the quality of translation testing. The authors asked translation students to translate an excerpt of a newspaper article taken

from *The Guardian*. In the translation brief, it was stated that a Dutch translation of the excerpt was to be published in the science section of a quality newspaper. The English source text reports on a study of the metal composition of a dagger found in the wrapping of the mummified King Tutankhamon (cf. Appendix). Two translation lecturers of KU Leuven and Zuyd University of Applied Sciences (cf. *infra*) independently picked this text out of a selection of five English newspaper articles on recent scientific developments that were thought to resemble the popular science articles that are regularly used for translation assignments in third-year BA translation courses in these programmes.

The translation students ($n = 57$) were given the newspaper article in full on paper and were asked to translate an excerpt of 268 words. They were given 90 minutes to complete the test. A pretest among four students from the two translation programmes had proved this time slot to be sufficient. The students produced their translations in Word and were allowed to consult electronic dictionaries and all resources available on the internet.

To study the reliability (and the validity) of the PIE method among a heterogeneous population of translation students, data were collected from translation students from different translation programmes and at different stages of translator training: 31 third-year students from the BA in University of Antwerp, 13 third-year students from the BA in Zuyd University of Applied Sciences (Vertaalacademie) and 13 students from the MA in Translation from KU Leuven participated in the experiment.

The three translation programmes have slightly different profiles: whereas the translation programmes offered at University of Antwerp and KU Leuven are both academic programmes, Zuyd University of Applied Sciences offers vocational training. The BA in University of Antwerp is a three-year academic study programme, in which two foreign languages and Dutch are studied as well as a specialisation track in translation that starts in the second year. The BA in Zuyd University of Applied Sciences is a four-year higher vocational programme, which offers specialized translation training in English, a foreign language of choice and Dutch. The MA in and KU Leuven is a one-year academic programme focused on translation from two foreign languages into Dutch. In other words, the students from the three study programmes were not only at different stages in translator training, but they also had different levels of translation experience.

Prior to the evaluation, selections were made of items that were considered appropriate.¹² During the selection process by far the most attention was devoted to cultural and linguistic differences (English vs. Dutch) and to the prospective TT function (that was mentioned in the brief). Shortly after the selection process, the correct translation solutions that readily sprang to mind were noted down. Below the reader will find the two independent selections that have been made and the lists of correct solutions. The selection under the heading of University of Antwerp (Table 1) originally comprised 31 items, but the number was brought back to 28, due to technological constraints.¹³ The selection under the heading of KU Leuven and Zuyd (Table 2) originally consisted of 25 items, but now contains no more than 20 items.¹⁴ What is immediately apparent, is that, despite some differences the teams of selectors, who have made an independent selection, seemed to be more or less on the same page with regard to the selected items. Differences that do leap to the eye (e.g. “contamination” in the University of Antwerp selection does not appear in The KU Leuven and Zuyd selection, and “decades” in the KU Leuven and Zuyd selection does not appear in the University of Antwerp selection), however, cannot be explained away by referring to the students’ or institution’s profile – a factor that is, nonetheless, bound to influence the selection process. Furthermore, when trying to identify the differences between the preselections of correct solutions, one is inclined to say that there is no striking disparity between the two. The similarities seem to suggest that a clear translation brief and a certain level of teaching and evaluation

How to Put the Translation Test to the Test?

Table 1. Set of test items selected by Selectors UA

Items	Correct Solutions
(1) made with	gemaakt van/uit/in vervaardigd van/uit/in
(2) Thursday 2 June 2016 07.10 BST	Ø donderdag 2 juni 2016
(3) Suggests	wijst op aanwijzingen aanwijzing
(4) Wrapping	Windsels windselen doeken
(5) Burial chamber	Grafkamer
(6) Urn	Kist
(7) Remains	stoffelijk overschot lichaam stoffelijke resten
(8) Ancient	Eeuwenoude
(9) Contamination	schadelijke stoffen aantasting en licht andere schadelijke elementen
(10) King Tutankhamun	koning Toetanchamon farao Toetanchamon
(11) in 1925,	In 1925 [no comma]
(12) 3,300 years	3300 jaar
(13) along with its levels of cobalt	in combinatie met/samen met/evenals het kobaltgehalte samen met de hoeveelheid kobalt in combinatie met sporen van kobalt en de aanwezigheid van kobalt samen met de hoeveelheid aan kobalt in combinatie met kobalt alsook/ook kobalt samen met een bepaald gehalte aan kobalt
(14) rock crystal pommel	pommel van/uit rotskristal pommel van/uit bergkristal degenknop van/uit rotskristal degenknop van/uit bergkristal
(15) decorated sheath	versierde schede schede versierd met
(16) Carter's	Carters van Carter
(17) in ancient Egypt	in het oude Egypte in het klassieke Egypte
(18) An x-ray fluorescence spectrometer	Röntgenfluorescentiespectrometer röntgenfluorescentiespectrometrie
(19) Known	Bekende Gevonden
(20) 2,000	2000 Tweeduizend
(21) [2,000]km	Km
(22) the Red Sea coast	de Rode Zeekust

continued on following page

Table 1. Continued

Items	Correct Solutions
(23) 150 miles	Ø
(24) Alexandria	Alexandrië
(25) Alexander the Great	Alexander de Grote
(26) Seaport city	havenstad zeehaven van
(27) Was known as	bekendstaan als bekend was als genoemd werd
(28) The Fourth century BC	de vierde eeuw voor Christus de vierde eeuw v.C. de vierde eeuw v.Chr.

experience effectively do provide us with the hoped-for operational measures for TT evaluation and, as a consequence, do add to the construct validity of PIE testing.

Having inputted the source text, the selections of items and all student versions in the evaluation module of the abovementioned software translationQ, the program first automatically identified the predicted correct translations for every item in both lists. At that point, the evaluators were given a rough idea of what perturbative theory calls “the degree of approximation”. The provisional percentages, which are displayed in translationQ, answered the question: How valuable is the provided “exact solution” in this context? By automatically identifying correct items and listing every correctly identified solution in the student versions, translationQ steers evaluators into the direction of possible unforeseen correct solutions and of incorrect solutions. Those solutions were scrutinized by the evaluators, who continued to work independently of each other. The evaluators then added correct student translations to their respective lists (cf. Table 1 and Table 2). The addition of correct items, which can be aptly described as “small terms”, marked the first perturbative moment in this PIE test. The reader will note that part of the differences between preselected solutions were smoothed out as a result of this perturbative process. However, the definitive selection of items and solutions still renders visible the hand of the evaluator in the evaluation process. For instance, “gemaakt met” does not appear in University of Antwerp’s list of correct solutions. There can be various reasons for discrepancies. Most often, correction is governed by or backed up with language rules that have been formulated in authoritative sources. In the case of “gemaakt met”, a reference to authoritative sources like the Dutch Language Union (Taalunie) and Van Dale suffices to explain some of the decisions made by University of Antwerp evaluators. However, sometimes the rules are simply not that strict or have relaxed with time. In some cases, no rule can be applied. In these situations, the evaluators call the shots.¹⁵ Still, blank spots and grey areas seem inevitable in translation evaluation. Therefore, it is mandatory that the inter-rater and intra-rater reliability of test designs in translation evaluation are guaranteed: reliable testing prevents the (aspiring) translators from being duped by these blank spots and grey areas.

As soon as the added solutions were inputted, translationQ recommenced the identification of correct translation solutions and recalculated the PIE score for all 57 students. At this stage, the evaluators assessed the appropriateness of the items. By pressing the “PIE Export” button in the tool, a multitable Excel-report of the PIE scores was generated and flanked by docimological data. The docimological

How to Put the Translation Test to the Test?

Table 2. Set of test items selected by selectors of KU Leuven and Zuyd

Items	Correct Answer/s
(1) Made with	vervaardigd van/uit/met gemaakt van/uit/met
(2) Thursday 2 June 2016 07.10 BST	∅ donderdag 2 juni 2016 Donderdag 2 juni 2016
(3) Metal composition	metalen samenstelling metaalsamenstelling materiaalsamenstelling
(4) Suggests	wijst op aanwijzing aanwijzingen
(5) Origin	Oorsprong Herkomst
(6) Ancient	Eeuwenoude uit de oudheid
(7) Dagger entombed with	dolk die samen met farao/koning Toetanchamon begraven werd/werd/is begraven dolk met farao/koning Toetanchamon begraven werd/werd/is begraven dolk die begraven werd/is/werd begraven dolk uit farao/koning Toetanchamons graftombe dolk die samen met koning Toetanchamon in de tombe werd begraven
(8) Archeologist Howard Carter	archeoloog Howard Carter
(9) 3,300	3300 3.300 meer dan 3000/drieduizend
(10) Lily and jackal-decorated sheath	een schede versierd met lelies en jakhalzen een met lelies en jakhalzen versierde schede een versierde schede
(11) Has puzzled	stelde/zette [...] voor een raadsel verbaasd over liet/deed [...] perplex staan het hoofd [...] gebroken
(12) the decades	decennialang de decennia tientallen Jaren
(13) Carter's	Carters van Carter
(14) x-ray fluorescence spectrometer	een röntgenfluorescentiespectrometer röntgenfluorescentiespectrometrie
(15) 2,000km	2000 kilometer 2.000 kilometer 2000 km 2.000 km
(16) Red Sea coast of Egypt	de kust van de Rode Zee in Egypte Egyptische kust van de Rode Zee
(17) Levels	Gehalte gehalten/gehaltes concentraties waarden nikkel- en kobaltgehaltes samenstelling
(18) 150 miles (240km)	240 kilometer (omit: 150 miles) tweehonderdveertig kilometer (omit: 150 miles)
(19) West of	ten westen van
(20) The seaport city of	∅ de havenstad

How to Put the Translation Test to the Test?

data are particularly relevant in the perturbative context. The data of the PIE test are presented in Tables 3 and 4.

In the context in which the present experiment was carried out, it was paramount that items were not found to be too difficult or too easy. The p value is a fraction of the sum total of items and is displayed as a number between 0 and 1. For an item to be up to the mark, the p value should lie between 0.27 and 0.79. As the p value is indicative of item difficulty, it is recommended that items with diverging values be removed from the list. As can be seen below, it is not uncommon that the mark is undershot or overshot (items with an acceptable p value have been shaded in Tables 3 and 4). In Table 3, 15 items in University of Antwerp's list fail to meet the quality expectations. In Table 4, 5 items in the KU Leuven and Zuyd Sciences list do not come up to standard.

Table 3. P values and D indices of University of Antwerp Items

Items	P Value	d Index
(1) made with	0.93	0.25
(2) Thursday 2 June 2016 07.10 BST	0.75	0.06
(3) Suggests	0.47	0.31
(4) Wrapping	0.26	0.19
(5) Burial chamber	0.84	0.31
(6) Urn	0.06	0.07
(7) Remains	0.18	0.00
(8) Ancient	0.23	0.19
(9) Contamination	0.30	-0.25
(10) King Tutankhamun	0.89	0.06
(11) in 1925,	0.98	0.00
(12) 3.300 years	0.56	-0.19
(13) along with its levels of cobalt	0.25	0.25
(14) rock crystal pommel	0.32	0.19
(15) decorated sheath	0.42	0.44
(16) Carter's	0.63	0.13
(17) in ancient Egypt	0.89	0.00
(18) An x-ray fluorescence spectrometer	0.44	0.50
(19) Known	0.93	0.19
(20) 2,000	0.75	0.56
(21) [2,000]km	0.75	0.56
(22) the Red Sea coast	0.71	0.38
(23) 150 miles	0.86	0.25
(24) Alexandria	0.81	0.38
(25) Alexander the Great	0.96	0.06
(26) Seaport city	0.70	0.06
(27) Was known as	0.04	0.00
(28) The Fourth century BC	0.53	0.38

How to Put the Translation Test to the Test?

In order to complete the next stage in the perturbative process, the p values should be interpreted in tandem with the d indices. As mentioned, d indices shed light on the discriminatory power of items. In other words, an item's d index provides an answer to the question: How likely is it that this item will allow for distinction between good and bad translations? The aptness of an item is attested to when the d index of an item with a satisfying p value is higher than 0.29 (items with an acceptable d index have been shaded in Tables 3 and 4). As can be inferred from Tables 3 and 4, only 7 items in the University of Antwerp list and 6 items in the KU Leuven and Zuyd list have an acceptable difficulty level as well as sufficient discriminatory power (in both tables, the docimologically justified items have been marked (in bold)).

It should be recalled that, ideally, only the items with acceptable p values and acceptable d indices will make the cut. This means that the docimological data have a serious impact on the selections and, consequently, on their respective critical mass. Experience has taught that the reduction of critical mass need not be a reason for concern. The docimological data are seldom uplifting. This is all the more true in a training setting, where the number of trainees is often limited and where competence levels of examinees are often more or less on a par. Still, the data shed crucial light on the quality of the test (design). The question that now begs an answer is: How many acceptable items are needed to reach a critical mass, and, as a consequence, to warrant the design of a translation test? An easy answer to this question cannot be provided. At present, there is no way of telling how many docimologically justified

Table 4. P values and D indices of KU Leuven and Zuyd University of Applied Sciences items

Items	P Value	d Index
(1) Made with	0.96	0.00
(2) Thursday 2 June 2016 07.10 BST	0.75	0.06
(3) Metal composition	0.42	0.50
(4) Suggests	0.47	0.31
(5) Origin	0.49	0.06
(6) Ancient	0.23	-0.06
(7) Dagger entombed with	0.49	0.20
(8) Archeologist Howard Carter	0.86	0.31
(9) 3.300	0.96	0.13
(10) Lily and jackal-decorated sheath	0.28	0.69
(11) Has puzzled	0.40	0.13
(12) the decades	0.49	0.44
(13) Carter's	0.63	0.13
(14) x-ray fluorescence spectrometer	0.28	0.38
(15) 2,000km	0.75	0.13
(16) Red Sea coast of Egypt	0.63	0.19
(17) Levels	0.60	0.31
(18) 150 miles (240km)	0.77	0.19
(19) West of	0.86	0.06
(20) The seaport city of	0.67	0.19

How to Put the Translation Test to the Test?

items will suffice for the (PIE) testing of target text quality. Research needs to be carried out on this topic and/or agreement is to be reached.

As mentioned above, the Excel-report of translationQ also contained the raw scores as well as the recalculated PIE scores of all 57 students. The raw score is always a fraction of the total number of items in the original selection (in the test cases: a fraction of 28 and 20). The recalculated score ignores docimologically unjustified items. In other words, the recalculated PIE score is always a fraction of the number of items that were found appropriate (in the test cases: a fraction of 7 and 6).

Table 5 shows the students' individual scores on each test. As is evident from the percentages that flank the student scores, the differences between the raw scores and the recalculated scores are at times dramatic.¹⁶ For instance, there is a large difference between the raw scores and the recalculated scores of

Table 5. Individual scores of the candidates on the PIE test

Student No.	University of Antwerp: "Raw" Pie Score (/28)	University of Antwerp: Recalculated Pie Score (/7)	Ku Leuven and Zuyd University of Applied Sciences: "Raw" Pie Score (/20)	Ku Leuven and Zuyd University of Applied Sciences: Recalculated Pie Score (/6)	Scores: University of Antwerp, Ku Leuven and Zuyd University of Applied Sciences Combined (/10)
1	12 (43%)	4 (57%)	6 (30%)	1 (17%)	3 (30%)
2	12 (43%)	1 (14%)	7 (35%)	1 (17%)	1 (10%)
3	14 (50%)	4 (57%)	7 (35%)	3 (50%)	5 (50%)
4	16 (57%)	2 (29%)	8 (40%)	1 (17%)	2 (20%)
5	13 (46%)	1 (14%)	8 (40%)	1 (17%)	2 (20%)
6	14 (50%)	3 (43%)	7 (35%)	1 (17%)	4 (40%)
7	14 (50%)	2 (29%)	11 (55%)	1 (17%)	3 (30%)
8	15 (54%)	4 (57%)	9 (45%)	2 (33%)	4 (40%)
9	12 (43%)	1 (14%)	10 (50%)	1 (17%)	2 (20%)
10	14 (50%)	2 (29%)	8 (40%)	2 (33%)	3 (30%)
11	17 (61%)	2 (29%)	9 (45%)	1 (17%)	3 (30%)
12	13 (46%)	2 (29%)	9 (45%)	2 (33%)	4 (40%)
13	15 (54%)	1 (14%)	9 (45%)	1 (17%)	2 (20%)
14	13 (46%)	1 (14%)	8 (40%)	2 (33%)	3 (30%)
15	16 (57%)	4 (57%)	10 (50%)	1 (17%)	4 (40%)
16	16 (57%)	4 (57%)	11 (55%)	1 (17%)	4 (40%)
17	16 (57%)	5 (71%)	10 (50%)	2 (33%)	5 (50%)
18	15 (54%)	6 (86%)	11 (55%)	3 (50%)	6 (60%)
19	14 (50%)	1 (14%)	10 (50%)	3 (50%)	3 (30%)
20	14 (50%)	2 (29%)	11 (55%)	2 (33%)	2 (20%)
21	16 (57%)	4 (57%)	9 (45%)	2 (33%)	6 (60%)
22	15 (54%)	3 (43%)	9 (45%)	2 (33%)	4 (40%)
23	15 (54%)	3 (43%)	10 (50%)	3 (50%)	5 (50%)
24	16 (57%)	4 (57%)	10 (50%)	1 (17%)	4 (40%)
25	16 (57%)	4 (57%)	9 (45%)	2 (33%)	4 (40%)
26	15 (54%)	6 (86%)	12 (60%)	1 (17%)	6 (60%)
27	15 (54%)	4 (57%)	9 (45%)	1 (17%)	5 (50%)

continued on following page

How to Put the Translation Test to the Test?

Table 5. Continued

Student No.	University of Antwerp: "Raw" Pie Score (/28)	University of Antwerp: Recalculated Pie Score (/7)	Ku Leuven and Zuyd University of Applied Sciences: "Raw" Pie Score (/20)	Ku Leuven and Zuyd University of Applied Sciences: Recalculated Pie Score (/6)	Scores: University of Antwerp, Ku Leuven and Zuyd University of Applied Sciences Combined (/10)
28	19 (68%)	5 (71%)	11 (55%)	2 (33%)	6 (60%)
29	16 (57%)	3 (43%)	12 (60%)	3 (50%)	4 (40%)
30	15 (54%)	3 (43%)	11 (55%)	4 (67%)	5 (50%)
31	15 (54%)	4 (57%)	12 (60%)	2 (33%)	6 (60%)
32	16 (57%)	3 (43%)	10 (50%)	3 (50%)	6 (60%)
33	17 (61%)	4 (57%)	12 (60%)	3 (50%)	4 (40%)
34	17 (61%)	3 (43%)	11 (55%)	1 (17%)	3 (30%)
35	17 (61%)	3 (43%)	11 (55%)	3 (50%)	6 (60%)
36	17 (61%)	5 (71%)	13 (65%)	4 (67%)	8 (80%)
37	14 (50%)	4 (57%)	12 (60%)	1 (17%)	5 (50%)
38	18 (64%)	6 (86%)	14 (70%)	3 (50%)	7 (70%)
39	19 (68%)	6 (86%)	12 (60%)	4 (67%)	9 (90%)
40	18 (64%)	7 (100%)	13 (65%)	3 (50%)	8 (80%)
41	17 (61%)	6 (86%)	12 (60%)	1 (17%)	6 (60%)
42	18 (64%)	7 (100%)	14 (70%)	4 (67%)	9 (90%)
43	17 (61%)	4 (57%)	12 (60%)	2 (33%)	4 (40%)
44	20 (71%)	4 (57%)	12 (60%)	5 (83%)	8 (80%)
45	20 (71%)	5 (71%)	12 (60%)	4 (67%)	7 (70%)
46	17 (61%)	5 (71%)	13 (65%)	3 (50%)	7 (70%)
47	18 (64%)	6 (86%)	12 (60%)	3 (50%)	6 (60%)
48	20 (71%)	6 (86%)	13 (65%)	4 (67%)	7 (70%)
49	18 (64%)	6 (86%)	13 (65%)	5 (83%)	10 (100%)
50	19 (68%)	5 (71%)	11 (55%)	4 (67%)	9 (90%)
51	18 (64%)	5 (71%)	13 (65%)	4 (67%)	8 (80%)
52	19 (68%)	6 (86%)	12 (60%)	4 (67%)	8 (80%)
53	22 (79%)	6 (86%)	15 (75%)	5 (83%)	8 (80%)
54	20 (71%)	4 (57%)	12 (60%)	4 (67%)	7 (70%)
55	22 (79%)	6 (86%)	13 (65%)	4 (67%)	8 (80%)
56	20 (71%)	5 (71%)	14 (70%)	4 (67%)	8 (80%)
57	22 (79%)	7 (100%)	17 (85%)	5 (83%)	9 (90%)

candidate 13 and candidate 40 and 42 on the University of Antwerp tests: whereas the score of candidate 13 comes crashing down after recalculation (54% vs. 14%), the scores of candidates 40 and 42 – at first but a meagre 64% – are skyrocketing up to 100%. In the KU Leuven and Zuyd test the differences are no less marked: the fairly acceptable raw scores of candidates 37 and 41 (60%), for example, plummet all the way to 17% after recalculation; it is remarkable that candidate 44 is one of the few students to see his/her odds of passing the test improve through recalculation (60% vs. 83%). Still, it should be noted that, although the percentages in the table might suggest otherwise, the PIE score need not be automatically converted into a grade. The benchmarks for passing and failing are to be set by the evaluator(s). It

must be borne in mind that the docimological data are brought into play to test or attest to the quality of the test (design).

What the results in Table 5 seem to suggest is that the critical mass of the remaining items in both selections offer relatively meagre data upon which to work and to pass an evaluative judgment. As mentioned in the previous section, the warrantability and accuracy of a PIE test need not be at stake because of a relative lack of critical mass in this phase of testing. One more step in the perturbative process can be taken: the evaluator(s) may see it fit to replace (some of the) flawed items in the selection. This is what the authors have decided to do. The most convenient way to take the edge off the recalculated scores in the 3rd and 5th columns would be to combine the docimologically justified items of both lists and leave out the doubles. It is needless to say that the combination of University of Antwerp, KU Leuven and Zuyd items is only a stopgap solution: in a normal test situation, there would only be one selection of items. However, this solution is advanced in this chapter to account for the possibility of adding/replacing acceptable items. The scores that are presented in the 6th column are based on the responses of the students to the following 10 test items: 1) suggest, 2) x-ray fluorescence spectrometer, 3) 2,000, 4) 2,000km, 5) Red Sea coast, 6) the fourth century BC, 7) metal composition, 8) lily and jackal-decorated sheath, 9) decades, 10) levels. By combining the item sets, the critical mass is increased, the differences in recalculated scores seem somewhat smoothed out, and test quality is more likely to be up to standard.¹⁷

CONCLUSION

In the introduction to this chapter, a backdrop has been provided which allowed the authors to sketch the contours of translation quality. In recent years, it has become abundantly clear that there is no single touchstone for reliable and valid testing of translation quality in all its complexity. The tools and methods at hand can only seek and serve to gauge particular aspects of translation quality. These limitations become all the more apparent, when attention is drawn to the broadness of the notion of translation itself. In consequence of this observation, the concept of “translation quality” can be said to cover the TT (product), the translation process proper (the trajectory from ST to TT) and the translation service (from project preparation all the way to post-production activities); and it can even be linked up, albeit in a circuitous manner, with translator competence.

Founded on the assumption that every tool and every method that serves to measure translation quality is more than likely to have obvious limitations, the PIE method has been presented as a method to evaluate (and assess) the quality of translation products. By showing how the method is imbricated in the history of translation quality research, presenting it as an analytical evaluation method, touching upon its compatibility with category-based analytical methods, and establishing its similarity to rich points evaluation as well as CDI evaluation, the authors have uncovered the reasons for the development of the PIE method. In the opinion of the authors, the efforts to overcome problems related to reliability and validity have thus far been too few and too ineffective.

Although PIE is still conceived of as an analytical method and will never be touted as the cure-all to all problems in translation quality testing, it is believed to have an advantage over the criteria-based and norm-referenced methods that have been reviewed in this chapter. The main strengths of the PIE method are listed below, alongside its weaknesses:

Strengths:

How to Put the Translation Test to the Test?

1. PIE saves time;
2. PIE evaluates candidates on the same items;
3. PIE is committed to a binary logic (no grey zone);
4. PIE guarantees inter-rater reliability;
5. PIE guarantees intra-rater reliability;
6. PIE allows for docimological item justification;

PIE allows for item selection on the basis of desired learning outcomes/competences (construct and content validity). Weaknesses:

1. PIE does not do away with subjective influences in the selection phase;
2. PIE does not do away with subjective influences in the correction phase;

PIE (usually) does not account for the text as a whole. As evidenced in this chapter, various reasons can be adduced in favour of using PIE. PIE tests are usually very time efficient. Contrary to the vast majority of analytical methods, PIE guarantees that each candidate will be evaluated on the same items, as every PIE test is based on a fixed selection of items. In contrast to most analytical methods, PIE assures that every candidate is treated equally in the correction process: first of all, PIE is committed to a binary logic according to which an answer is either correct or it is not (and this logic applies to all translations); secondly, weighting is avoided throughout the evaluation process. Taken together, these arguments provide credibility for the authors' belief that reliability, especially intra-rater and inter-rater reliability, is increased through PIE testing. All the more so, because, in the evaluation process, PIE allows evaluators to keep tabs on the quality of the selected items: docimological data show how valuable a test item is. Reliability can also be safeguarded in CDI evaluation, which also makes use of docimological observations. However, in stark contrast to CDI evaluation, PIE testing still grants the evaluator an important role in the selection process. Although the ultimate goal seems to be to rid translation of subjective influences, the somewhat undesired presence of evaluators does have one advantage: evaluators exert control over the validity of PIE tests (e.g. construct and content validity).

By bringing together the main strengths of criteria-based and norm-referenced evaluation, the authors believe to have found a testing technique that allows for better tailoring of translation product testing to the test situation. How this tailoring process works has been illustrated with an analogy (set up) between PIE and perturbation theory. As in perturbative approaches, a PIE test relies heavily on surefire solutions to returning problems in translation testing, solutions that have proved their worth in the past. In the case of translation testing, these solutions are formulated by the assessors/evaluators, who have accumulated knowledge and experience. However, on the basis of the test (e.g. versions) and with the aid of docimological data, elements and items can be added to and removed from the solution. The removal and the addition of elements and items are key in the perturbative process. By carefully following all the steps of PIE testing, one can get a firmer handle on the quality of translation product tests, and, most probably, on the quality of the tested TTs.

Still, it should be noted that there probably lies some uncharted territory between where we are in translation testing, and where we need to be. If PIE is to cement its reputation as a reliable and valid testing technique, some methodological improvements and clarifications are still required. A number of pertinent questions that may inform future research are: Which criteria should be applied in item preselection? Can/Should the preselection process be systematized so as to reduce the influence of the

evaluator to a minimum? How many evaluators are needed to obtain a valid selection of items? What level of expertise is required for the selection of items and the evaluation of given solutions? How many evaluators are needed to reach a sound judgment on the correctness of given solutions? Would it be possible to automate the evaluation of solutions? How many test translations are required to reach a point of docimological saturation? What number of docimologically justified items is required to reach a critical mass? Can the selection of weak items be prevented in any way? Would it be beneficial to add other docimological yardsticks to PIE analysis? Does the reliability and/or validity of PIE increase when it is undergirded by analytical grids? How does PIE correlate with other methods presented in this chapter? What are the ergonomic qualities of PIE in terms of cognitive effort and time management? Despite all the progress that has been made, all these questions, and many more, fairly cry for answers.

ACKNOWLEDGMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

- Alves, F., & Hurtado-Albir, A. (2017). Perspectives for research on cognitive aspects of translation. In J. W. Schwieter & A. Ferreira (Eds.), *The handbook of translation and cognition* (pp. 537–553). Oxford, UK: Wiley Blackwell. doi:10.1002/9781119241485.ch29
- Anckaert, P., & Eyckmans, J. (2007). IJken en ijken is twee. Naar een normgerelateerde ijkpuntenmethode om vertaalvaardigheid te evalueren. In C. Van de Poel & W. Segers (Eds.), *Vertalingen objectief evalueren: Matrices en Ijkpunten* (pp. 53–81). Leuven: Acco.
- Anckaert, P., Eyckmans, J., & Segers, W. (2006). Vertaalvaardigheid evalueren: Een normgerelateerde benadering. *Tijdschrift n/f*, 6, 9–27.
- Anckaert, P., Eyckmans, J., & Segers, W. (2008). Pour une évaluation normative de la compétence de traduction. *ITL: Review of Applied Linguistics*, 41(155), 53–76. doi:10.2143/ITL.155.0.2032360
- Angelelli, C., & Jacobson, H. E. (Eds.). (2009). *Testing and assessment in translation and interpreting studies*. Amsterdam: John Benjamins. doi:10.1075/ata.xiv
- Beeby, A. (2000). Evaluating the Development of Translation Competence. In C. Schäffner & B. Adab (Eds.), *Developing translation competence* (pp. 185–198). Amsterdam: John Benjamins. doi:10.1075/btl.38.18bee
- Beeby, A., Fernández, M., Fox, O., Albir, A., Kozlova, I., Kuznik, A., ... Wimmer, S. PACTE. (2009). Results of the validation of the PACTE translation competence model: Acceptability and decision making. *Across Languages and Cultures*, 10(2), 207–230. doi:10.1556/Acr.10.2009.2.3
- Beeby, A., & Martínez, N. (1998). *La evaluación en la enseñanza universitaria de la traducción profesional: Un estudio empírico*. Paper presented at the EST Congress, Granada.

How to Put the Translation Test to the Test?

Bloom, B. S. (1972). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive Domain*. New York: David McKay.

Bowker, L. (2001). Towards a methodology for a corpus-based approach to translation evaluation. *Meta: Journal des Traducteurs/Meta: Translators' Journal*, 46(2), 345–364.

Castillo Rincón, L. M. (2010). *La evaluación de la aceptabilidad de las traducciones: Un estudio exploratorio sobre la idoneidad de los puntos ricos para evaluar traducciones* (Master's thesis). Universitat Autònoma de Barcelona, Barcelona, Spain. Retrieved June 22, 2017, from <http://hdl.handle.net/2072/179692>

Castillo Rincón, L. M. (2015). Acquisition of translation competence and translation acceptability: An experimental study. *The International Journal for Translation & Interpreting Research*, 7(1), 72–85.

Chesterman, A. (1997). *Memes of translation: The spread of ideas in translation theory*. Amsterdam: John Benjamins. doi:10.1075/btl.22

Colina, S. (2009). Further evidence for a functionalist approach to translation quality evaluation. *Target: International Journal on Translation Studies*, 21(2), 235–264. doi:10.1075/target.21.2.02col

Delisle, J. (1980). *L'analyse du discours comme méthode de traduction: Théorie et pratique*. Ottawa: University of Ottawa.

Delisle, J. (1990). *L'enseignement de l'interprétation et de la traduction: De la théorie à la pédagogie*. Ottawa: University of Ottawa.

Delisle, J. (1993). *La traduction raisonnée: Manuel d'initiation à la traduction professionnelle de l'anglais vers le français*. Ottawa: University of Ottawa.

Delisle, J. (1998). Définition, rédaction et utilité des objectifs d'apprentissage en enseignement de la traduction. In I. I. Garcia & J. Verdegel (Eds.), *Los estudios de traducción: Un reto didáctico* (pp. 13–44). Castellón: Universitat Jaume I.

Drugan, J. (2013). *Quality in professional translation assessment and improvement*. London: Bloomsbury.

Ebel, R. L. (1954). Procedures for the analysis of classroom tests. *Educational and Psychological Measurement*, 1(2), 352–364. doi:10.1177/001316445401400215

EMT Expert Group. (2009). *Competences for professional translators, experts in multilingual and multimedia communication*. Retrieved January, 8, 2018, from https://ec.europa.eu/info/sites/info/files/emt_competences_translators_en.pdf

Even-Zohar, I. (1990). The position of translated literature within the literary polysystem. *Poetics Today*, 11(1), 45–51. doi:10.2307/1772668

Eyckmans, J., Anckaert, P., & Segers, W. (2009). The perks of norm-referenced translation evaluation. In C. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in Translation and Interpreting Studies* (pp. 73–93). Amsterdam: John Benjamins. doi:10.1075/ata.xiv.06eyc

- Eyckmans, J., Anckaert, P., & Segers, W. (2016). Translation and interpretation skills. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 219–235). Boston: Walter De Gruyter. doi:10.1515/9781614513827-016
- Fields, P., Hague, D., Koby, G., Lommel, A., & Melby, A. K. (2014). What is quality? A management discipline and the translation industry get acquainted. *Revista Tradumàtica: tecnologies de la traducció*, 12, 404–412.
- Garant, M. (2009). A case for holistic translation assessment. *AFinLA-e: Soveltavan kielitieteen tutkimuksia*, 1, 5–17.
- Garvin, D. G. (1984). What does product quality really mean? *Sloan Management Review*, 26(1), 25–43.
- Garvin, D. G. (1987). Competing on the eight dimensions of quality. *Harvard Business Review*, 65(6), 101–109.
- Görög, A. (2014). Quantification and comparative evaluation of quality: The TAUS dynamic quality framework. *Revista Tradumàtica*, 12, 443–454. doi:10.5565/rev/tradumatica.66
- Greene, B. (2011). *The elegant universe: Superstrings, hidden dimensions, and the quest for the ultimate theory*. New York: Vintage Books.
- Hermans, T. (1991). Translation norms and correct translations. In K. van Leuven-Zwart & T. Naaijken (Eds.), *Translation studies: The state of the art. Proceedings of the first James S. Holmes Symposium on Translation Studies* (pp. 155–169). Amsterdam: Rodopi.
- Holmes, J. S. (1988). *Translated! Papers on literary translation and translation studies*. Amsterdam: Rodopi.
- Hönig, H., & Kussmaul, P. (1982). *Strategie der Übersetzung: Ein Lehr- und Arbeitsbuch*. Tübingen: Narr.
- Hurtado-Albir, A. (2007). Competence-based curriculum design for training translators. *The Interpreter and Translator Trainer*, 1(2), 163–195. doi:10.1080/1750399X.2007.10798757
- Hurtado-Albir, A. (2017). *Researching translation competence by PACTE group*. Amsterdam: John Benjamins. doi:10.1075/btl.127
- International, S. A. E. (2005). *Translation quality metrics (J2450)*. Warrendale, PA: SAE.
- ISO. (2015). *ISO 17100 Translation services: Requirements for translation services*. Geneva: ISO.
- Klimkowski, K. (2016). Entrepreneurial aspects of translation competence in didactic reflection and practice. In Ł. Bogucki, B. Lewandowska-Tomaszczyk, & M. Thelen (Eds.), *Translation and Meaning: New series*, 2 (pp. 143–156). Frankfurt am Main: Peter Lang.
- Klimkowski, K. (2017). Empowering translation students in entrepreneurship: The concept of translation service provision competence in didactic application. In S. Hagemann, J. Neu, & S. Walter (Eds.), *Translationslehre und Bologna-Prozess: Unterwegs zwischen Einheit und Vielfalt* (pp. 189–213). Berlin: Frank & Timme.

How to Put the Translation Test to the Test?

- Koby, G., Fields, P., Hague, D., Lommel A., & Melby, A. K. (2014). Defining translation quality. *Revista Tradumàtica: Tecnologies de la traducció*, 12, 413–420.
- Koby, G. S., & Melby, A. K. (2013). Certification and job task analysis (JTA): Establishing validity of translator certification examinations. *The International Journal of Translation and Interpreting Research*, 5(1), 174–210. doi:10.12807/ti.105201.2013.a10
- Kockaert, H., & Segers, W. (2012). L'assurance qualité des traductions: Items sélectionnés et évaluation assistée par ordinateur. *Meta: Journal des Traducteurs/Meta: Translators' Journal*, 57(1), 159–176.
- Kockaert, H., & Segers, W. (2014). Evaluation de la traduction: La méthode PIE (Preselected Items Evaluation). *Turjuman*, 23(2), 232–250.
- Kockaert, H., & Segers, W. (2017). Evaluation of legal translations: PIE method (Preselected Items Evaluation). *Journal of Specialised Translation*, 27, 148–163.
- Kostiainen, T. (2012, October). *Tools for quality assurance and translation memory maintenance*. Paper presented at ATA Annual Conference, San Diego, CA.
- Lei, P.-W., & Wu, Q. (2007). CTTITEM: SAS macro and SPSS syntax for classical item analysis. *Behavior Research Methods*, 39(3), 527–530. doi:10.3758/BF03193021 PMID:17958163
- Localization Industry Standards Association (2004). *LISA QA model 3.0. product documentation*. Féchy: LISA.
- Lommel, A. (Ed.). (2014). *Multidimensional quality metrics (MQM) definition*. Retrieved June 20, 2017 from <http://www.qt21.eu/mqm-definition/definition-2014-08-19.html>
- Lommel, A., Burchardt, A., & Uszkoreit, H. (2014). Multidimensional quality metrics MQM: A framework for declaring and describing translation quality metrics. *Tradumatica*, 12(12), 455–463. doi:10.5565/rev/tradumatica.77
- Mariana, V., Cox, T., & Melby, A. K. (2015). The multidimensional quality metrics (MQM) framework: A new framework for translation quality assessment. *The Journal of Specialised Translation*, 23, 137–161.
- Martínez-Melis, N., & Hurtado Albir, A. (2001). Assessment in translation studies: Research needs. *Meta, Journal des Traducteurs/Meta: Translators' Journal*, 46(2), 272–287.
- Massardo, I., Van der Meer, J., & Khalilov, M. (2016). *TAUS translation technology landscape report*. De Rijp: TAUS.
- Matlock-Hetzel, S. (1997, January). *Basic concepts in item and test analysis*. Paper presented at Annual meeting of the Southwest Educational Research Association, Austin, TX.
- McAlester, G. (2000). The evaluation of translation into a foreign language. In C. Schäffner & B. Adab (Eds.), *Developing translation competence* (pp. 229–241). Amsterdam: John Benjamins. doi:10.1075/btl.38.21mca
- Melby, A., Fields, P., Hague, D., Koby, G., & Lommel, A. (2014). Defining the landscape of translation. *Revista Tradumàtica: tecnologies de la traducció*, 12, 392–403.

- Murdock, J. (2014). Perturbation models. In M. Grinfeld (Ed.), *Mathematical tools for physicists* (pp. 411–449). Somerset, NJ: Wiley.
- Nord, C. (1988). *Textanalyse und Übersetzen: Theoretische Grundlagen, Methode und didaktische Anwendung einer übersetzungsrelevanten Textanalyse*. Heidelberg, Germany: Groos.
- Nord, C. (1994). It's tea-time in wonderland: 'Culture-markers' in fictional texts. In H. Pürschel (Ed.), *Intercultural communication* (pp. 523–538). Frankfurt: Peter Lang.
- Nord, C. (1997). *Translating as a purposeful activity: Functionalist approaches explained*. Manchester, UK: St. Jerome.
- Nord, C. (2002). *Lernziel: Professionelles Übersetzen: Spanish-Deutsch*. Wilhelmsfeld: Gottfried Egert.
- Nord, C. (2010). Functionalist approaches. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of Translation Studies, 1* (pp. 120–128). Amsterdam: John Benjamins. doi:10.1075/hts.1.fun1
- Nord, C. (2011a). *Funktionsgerechtigkeit und Loyalität: Theorie, Methode und Didaktik des funktionalen Übersetzens*. Berlin: Frank & Timme.
- Nord, C. (2011b). *Defining translation functions: The translation brief as a guideline for the trainee translator*. Retrieved June 20, 2017 from <http://www.scribd.com/doc/39090525/NORD>
- O'Brien, S. (2012). Towards a dynamic quality evaluation model for translation. *Journal of Specialized Translation, 17*, 55–77.
- PACTE. (2002). Exploratory tests in a study of translation competence. *Conference Interpretation and Translation, 4*(2), 41–69.
- PACTE. (2003). Building a translation competence model. In F. Alves (Ed.), *Triangulating translation: Perspectives in process oriented research* (pp. 43–66). Amsterdam: John Benjamins.
- PACTE. (2008). First results of a translation competence experiment: 'Knowledge of translation' and 'efficacy of the translation process'. In J. Kearns (Ed.), *Translator and interpreter training: Issues, methods and debates* (pp. 104–126). London: Continuum.
- PACTE. (2011). Results of the validation of the PACTE translation competence model: Translation problems and translation competence. In C. Alvstad, A. Hild, & E. Tiselius (Eds.), *Methods and strategies of process research: Integrative approaches in translation studies* (pp. 317–343). Amsterdam: John Benjamins.
- PACTE. (2016a, July). *Nivelación de competencias en la adquisición de la competencia traductora*. Paper presented at DidTRAD III. Congreso internacional sobre investigación en Didáctica de la traducción, Barcelona, Spain.
- PACTE. (2016b, September). *Competence levels in translation: Working towards a European framework*. Paper presented at the 8th EST Congress, Aarhus.
- Popovič, A. (1975). *Teória umeleckého prekladu*. Bratislava: Tatran.

How to Put the Translation Test to the Test?

- Popovič, A. (1976). *Dictionary for the analysis of literary translation*. Edmonton: University of Alberta.
- Popovič, A. (2006). *La scienza della traduzione: Aspetti metodologici: La comunicazione traduttiva* (B. Osimo & D. Laudani, Trans.). Milan: Hoepli.
- Saldanha, G., & O'Brien, S. (2014). *Research methodologies in translation studies*. London: Routledge.
- Schäffner, C., & Adab, B. (Eds.). (2000). *Developing translation competence*. Amsterdam: John Benjamins. doi:10.1075/btl.38
- Segers, W. (2012). Vertaalevaluatie: Waarover hebben we het eigenlijk? *Ad Rem. Tijdschrift voor zakelijke communicatie*, 26(6), 17–18.
- Segers, W., & Kockaert, H. (2016). Can subjectivity be avoided in translation evaluation? In *Translation and meaning: New series, 1* (pp. 69–78). Frankfurt am Main: Peter Lang.
- Tenner, A. R., & DeToro, I. J. (1992). *Total quality management: Three steps to continuous improvement*. Reading, MA: Addison-Wesley.
- Thelen, M. (2016a). Quality and objectivity of testing and assessing in translator training: Dilemmas and suggestions. In M. Thelen, G. van Egdom, D. Verbeeck, & B. Lewandowska-Tomaszczyk (Eds.), *Translation and meaning: New serie, 1* (pp. 79–98). Frankfurt am Main: Peter Lang. doi:10.3726/978-3-653-05899-4/17
- Thelen, M. (2016b). The practice-oriented translator training curriculum: An example. *Current Trends in Translation Teaching and Learning E*, 3, 163–200.
- Valero-Garcés, C., Vigier, F., & Phelan, M. (2014, October). *Holistic versus analytical assessment of legal translations*. Paper presented at QUALETRA Final Conference, Antwerp.
- Van de Poel, C., & Segers, W. (Eds.). (2007). *Vertalingen objectief evalueren: Matrices en IJkpunten*. Leuven: Acco.
- Vandepitte, S. (2013, March). *Teaching professional standards and quality assurance*. Paper presented at Optimale WP 6, Brussels.
- Waddington, C. (2001). Different methods of evaluating student translation: The question of validity. *Meta: Journal des Traducteur/Meta: Translators' Journal*, 46(2), 312–325.
- Waddington, C. (2004). Should translations be assessed holistically or through error analysis? *Lebende Sprachen*, 49(1), 28–35. doi:10.1515/LES.2004.28
- Wang, C. (2015, September). *Quantifying assessment criteria in translator training courses: A step forward for translator trainers and translator users to meet*. Paper presented at New perspectives in assessment in translation training: Bridging the gap between academic and professional assessment, London.
- Wiersma, W., & Jurs, S. G. (1990). *Educational measurement and testing*. Boston: Allyn & Bacon.
- Williams, M. (2009). Translation quality assurance. *Mutatis Mutandis*, 2(1), 3–23.

KEY TERMS AND DEFINITIONS

Analytical Evaluation: Evaluation method which analyses specific elements in translation products by means of an evaluation matrix which includes error categories and weighting.

Assessment: Process of measuring and understanding the effectiveness of teaching and learning strategies and methods, and provides information on how to improve learning and teaching.

Calibration of Dichotomous Items (or CDI): Method used in analytical evaluation which considers all source text elements as potential test items.

Evaluation: Assessment of a deliverable (a translation) or a person's knowledge and competence (a translator) against defined criteria. The value which is attached to a translation or translator is expressed as a score, a figure, or a letter code.

Holistic Evaluation (or Intuitive-Impressionistic Evaluation): Evaluation method which considers the translation products as a whole.

Inter-Rater Reliability: Degree to which two or more evaluators produce the same results in measuring the same characteristics. It establishes the relation between the scores of different evaluators: inter-rater reliability is high when a strong relation between the different scores is found.

Intra-Rater Reliability: Degree to which one and the same evaluator produces the same results in measuring the same characteristic. It establishes the relation between the scores of one and the same evaluator: intra-rater reliability is high when a strong relation between the different scores of the same evaluator is found.

Preselected Items Evaluation (or PIE): Analytical evaluation method to evaluate translation products on the basis of specific items from the source text which are selected on the basis of the evaluators' expertise, the translation brief, and the requirements of the test.

Reliability: An instrument is reliable when it is objective, consistent and reproducible, stable, and transparent.

Validity: An instrument is valid when it measures what it intends to measure, and nothing more.

ENDNOTES

¹ It should be noted that Popovič's dictionary has its redeeming features. For instance, Popovič does not go so far as to deny low-quality translations their status as "translations". The lemma "traduzi-one di bassa qualità"(low-quality translation) is found one page further into the dictionary (2006, p. 175).

² It is not superfluous to point out that, for the sake of intellectual honesty, the soft spots of PIE testing will never be downplayed in this article. One should never lose sight of authors' main trust: devising high-quality translation tests.

³ The updated version of the applied branch of Holmes's map, a version that has been drawn up by Munday (2001, p. 13), should be brought forward in support of this statement.

⁴ Another quality metric that dates back to the 1990's is the SAE J2450, developed by the Society of Automotive Engineers. The metric system was first presented in 1997 and amended by a European taskforce in 2001. This metric system is comprised of seven error categories (SAE International, 2005).

How to Put the Translation Test to the Test?

- ⁵ The list of analytical QA models is long. Amongst the most widely adopted analytical grids, one finds ATA's "Framework for Standardized Error Marking", QA Distiller, ITR Blackjack, the models developed by the MeLLANGE group... More information on some of these tools has been provided by Kostianen (2012) and Vandepitte (2013).
- ⁶ Many of the arguments against analytical evaluation had already been adduced by other scholars (cf. McAlester, 2000; Anckaert et al., 2008; Eyckmans, et al., 2009; Eyckmans et al., 2012; Anckaert et al., 2013; Valero-Garcés, Vigier, & Phelan, 2014; Kockaert & Segers, 2017)
- ⁷ To quote Nord's own words: "Instead of setting the standards for the production of an equivalent target text, it provides the basis for a comparison for the source text's offer of information with the target-text 'profile'" (Nord, 2010, p. 124).
- ⁸ With regard to time management, Beeby (2000, p. 185) and McAlester (2000, p. 231) note that holistic evaluation has been and, in some cases, remained the method of choice simply because marking criteria require a lot of attention and time.
- ⁹ In this recent article, Kockaert and Segers state that PIE Light has the arguable benefit of "the possibility to adopt PIE without p value and d index calculation" (2017, p. 153). By adopting this "light" approach, the added value of PIE is reduced, as PIE light is hardly distinguishable (if at all) from the rich points method.
- ¹⁰ The term "tentative theory" is employed by Popper in *Objective Knowledge* (1979) and is one of the four steps in his remarkably simple schema of scientific methodology (P1→TT→EE→P2) (p. 121) – a schema that has also been applied to the phenomenon of translation by none other than Andrew Chesterman (1997, pp. 16-17, pp. 117-123).
- ¹¹ Like every perturbative theory, a PIE test is inevitably exposed to the risk of what is technically called a "failure of (a) perturbation theory" (cf. Greene 2011, p. 289). When no approximate solution is reached, the quality of a translation test cannot be assured.
- ¹² It is not common usage to produce multiple item selections for one single test. However, the authors created two separate item sets to highlight the perturbative principles underlying PIE evaluation. In case of insufficient "critical mass", the internal validity of the test can be at stake. The evaluator can try to safeguard the internal validity by selecting extra items. By making a second selection, the authors were able to select extra items with demonstrably good docimological values.
- ¹³ The removed University of Antwerp items are: 1) researchers who analysed... an extraterrestrial origin'; 2) The sarcophagus of ... and other contamination; 3) Meteocritics and Planetary.
- ¹⁴ The removed KU Leuven and Zuyd items are: 1) '(...)' ; 2) strongly suggests an extraterrestrial origin (consistency in translation of same quotation); 3) [the entire sentence:] The sarcophagus of... and other contamination'; 4) one meteorite; 5) Meteocritics and Planetary.
- ¹⁵ As mentioned earlier in this chapter, subjective influences may be further minimized, or even neutralized, in the future with the aid of translation corpora, translation memories or error memories. An error memory can be developed in revisionQ, the revision module of translationQ.
- ¹⁶ The authors hasten to point out that the percentages have been calculated manually and that they have only been added to make the differences more salient.
- ¹⁷ Still, one must not shut one's eyes to the differences in the answer sheets of the selectors (cf. Tables 1 and 2). For the two item sets to be successfully merged, the acceptability of all answers ought to be renegotiated.

APPENDIX

Fragment of the source text and a number of selected items.

Dagger in Tutankhamun's tomb was made with iron from a meteorite

Researchers who analysed metal composition of dagger within wrapping of mummified teenage king say it 'strongly suggests an extraterrestrial origin'

[IMAGE]

The sarcophagus of King Tutankhamun in its burial chamber after the mummy was placed in a glass urn designed to protect the remains from humidity and other contamination. [...]

A dagger entombed with King Tutankhamun was made with iron from a meteorite, a new analysis on the metal composition shows. [...]

Source: <https://www.theguardian.com/world/2016/jun/01/dagger-king-tut-tomb-iron-meteorite-egypt-mummy>

Explanatory table:

made with = selected by selectors University of Antwerp, KU Leuven And Zuyd University Of Applied Sciences

wrapping = selected by selectors of University of Antwerp

dagger entombed with = selected by selectors KU Leuven And Zuyd University Of Applied Sciences

Researchers who analysed... = items that have been removed due to technological constraints

Chapter 3

Quality Assurance in Translation and Corporate Communications: Exploring an Interdisciplinary Interface

Gary Massey

Zurich University of Applied Sciences, Switzerland

Regine Wieder

Zurich University of Applied Sciences, Switzerland

ABSTRACT

While the nature and status of translators' work are changing due to technologisation and other factors, translation is acquiring a strategic function in organisations. The intercultural component of translation competence makes translators well positioned to play a key role in assuring quality in international corporate communications. But quality models envisage only restricted interactions between translators, clients and communications specialists. Moreover, evidence about translators' self-concepts shows them under-equipped to adopt the roles that meaningful cooperation with corporate communications suggests. This chapter reports on a pilot study at the interface between translation and corporate communications in Switzerland. Presenting findings from a survey of translation and communications professionals, it reveals underdeveloped feedforward and feedback cultures and a translator self-concept that underplays the mediatory, advisory added value of human translation. Concrete implications for quality assurance and translator education are drawn and future research is outlined.

INTRODUCTION

This chapter addresses the hitherto untapped potential that exists between corporate communications on the one hand, and translation on the other. As international communication management gains growing momentum as a profession and discipline in its own right and as translation acquires an increasingly strategic function in organisations (Huck-Sandhu, 2015, 2016; Massardo, van der Meer, & Khalilov,

DOI: 10.4018/978-1-5225-5225-3.ch003

2016), the intercultural mediation inherent in translators' work (Liddicoat, 2016) would seem to make them well positioned to play an integral part in assessing and assuring output quality in international, multilingual corporate communications. In view of the rapidly changing nature and status of translators' work due to technologisation and other factors, a closer integration of translation in international, multilingual corporate communications should present translators with welcome professional opportunities. Yet, dedicated quality standards for translation service provision and other quality-assurance and quality-assessment models envisage only restricted interactions by means of feedforward and feedback between translation professionals, client organisations and the communications specialists who work for them. The term feedforward is applied here not simply in the sense, habitually used in the media and communication community, of "information about receivers that is available to the sender prior to communicating with them and that can be used to improve the effectiveness of the communication" (Chandler & Munday 2011), but in the more general sense of "the modification or control of a process using its anticipated results or effects" (feedforward, 2017). The present authors' definition of feedback is similarly broad, namely "helpful information or criticism that is given to someone to say what can be done to improve a performance, product, etc." (feedback, 2017). Moreover, evidence of translators' professional self-concept (Katan, 2011, 2016), defined by Kiraly (1995, p. 100) as "the image of the translator's social role, the translator's appraisal of his or her competency for translating a particular text, and understanding of responsibility towards the other personalities in the translation context of situation (author, commissioner, user, and reader)", suggests that most translators are not yet equipped to adopt the evaluative, advisory and (co-)creative roles that meaningful cooperation with corporate communications specialists would suggest.

Against this background, the current chapter reports on a pilot study conducted by the authors at the interface between translation and corporate communications in Switzerland. After setting out the backdrop to the study and describing its motivation and design, the chapter presents the detailed results of a questionnaire survey of translators, translation project managers and communications specialists. It then proceeds to discuss the implications of the findings for translation quality and translator education, finishing up with a brief consideration of prospects for future research.

BACKGROUND

In the 21st century, the translation profession and those educating the professionals are confronted with multiple challenges. For some years, there have been unmistakable signs of the erosion of the professional status of translators (European Commission, 2012; Pym, Orrego-Carmona, & Torres-Simón, 2015) due to inadequate signalling, the spread of an amorphous paraprofessional market as well as qualification and certification issues. At the same time, technological advances have occasioned a noticeable shift in translators' principal tasks. Computer-assisted translation (CAT) systems have been supporting translators in their work for well over two decades, leading some to define translation as a specific form of human-computer interaction (O'Brien, 2012) involving the situated cognition of work done by the mind in conjunction with complex physical environments and socio-technical systems (cf. Risku, 2010, p. 103). One has only to consider the near-universal deployment of translation-memory (TM) systems in the language industry and the established market for post-editing (PE) machine-translation (MT) output to recognise how highly technologised the translator's workplace has become. In fact, PE has developed into a separate mode of translation in its own right, categorised as a value-added service in

the ISO 17100 (2015, p. 18) international standard for translation services, attracting a growing body of research (cf. Carl, Guterath, & Hansen-Schirra, 2015; Carl, Bangalore, & Schaeffer, 2016; Läubli, Fishel, Massey, Ehrensberger-Dow, & Volk, 2013; Gueberhof Arenas, 2009, 2014a, 2014b; Koponen, 2016; O'Brien, Winther Balling, Carl, Simard, & Specia, 2014) and taught in dedicated courses offered at translator-education institutions such as our own.

Translators can and do clearly benefit from the greater efficiency and productivity they are able to achieve through the deployment of assistive technologies such as TM and MT combined with PE. The advantages of their use in terms of efficiency and quality have been reported in numerous studies (e.g. Carl et al., 2015, 2016; Fiederer & O'Brien, 2009; Gueberhof Arenas, 2009, 2014a, 2014b). As Koponen (2016) points out in her survey of research into PE effort, the results are mixed, depending on the context of deployment. The same technologies that assist translators also seem to impose constraints, with some undesirable effects on translators' working conditions and practices as well as on the quality of their products. Research on translators' responses to technology (Olohan, 2011) and other aspects of organisational ergonomics (Ehrensberger-Dow & Massey, 2017) reveal a certain resentment of technologies imposed by clients or organisations and associated reward structures, especially for PE (Gueberhof Arenas, 2013). There are also indicators from cognitive and ergonomic translation processes research (TPR) that translators subject themselves to the constraints of the technology, which appear to impact negatively on their creativity and sense of autonomy (e.g. Ehrensberger-Dow & Massey, 2014, 2017) and their editing behaviour (Mellinger & Shreve, 2016). The limitations imposed by fast-spreading technologisation, the segmentation of translation workflows and the consequent shift in translators' tasks are also undermining translators' own conceptualisation of their role and responsibilities, leading to a self-perception, reflected in strong indicators from a large-scale survey, that they are members of a low-autonomy profession (LAP) (Katan, 2011, 2016). Increasing reliance on technology could encourage translators to deploy more routines and automaticity in their work. While superficially increasing productivity, this can lead to "fixation" (Dodds & Smith, 1999), identified by researchers as a barrier to knowledge integration, creative problem-solving and decision-making (e.g. Bayer-Hohenwarter, 2010, 2011; Prassl, 2010) and a possible contributor to developmental stagnation (e.g. Göpferich, 2012).

Deep learning (DL) and neural machine translation (NMT) are likely to change the translation profession even more profoundly. Research on NMT performance has produced comparatively positive results over the relatively short period of its deployment (cf. Wu, Schuster, Chen, Le, & Norouzi, 2016), and the Translation Automation Users Society (TAUS) predicts that "fully automatic useful translation" will become the generally accepted norm within the next two decades and MT will reach human translation quality levels (Massardo et al., 2016, p. 11). Moreover, a recent survey of leading machine learning researchers by the Universities of Oxford and Yale reveals that by 2032, artificial intelligence (AI) systems are predicted to be able to "translate a text written in a newly discovered language into English as well as a team of human experts [could]" (Grace, Salvatier, Dafoe, Zhang, & Evans, 2017, p. 14).

Ostensibly, the current status of human translation appears to be under threat, but economic research suggests that this will have an uneven impact, depending in large part on the type of tasks being performed. An employment study by the Federal Reserve Bank of St. Louis (Dvorkin, 2016) reveals stagnation in routine cognitive and routine manual employment levels between 1983 and 2014, but a steady rise in employment in non-routine cognitive jobs over the same period, suggesting that automation above all jeopardises routine work. In response to the employment polarisation underlying this trend, Canon and Marifian (2013, p. 11) recommend that "while educational achievement is undoubtedly important as demand continues to increase for college-educated, high-skilled (and high-wage) workers, it also may

be useful to emphasise development and training for non-routine skills since they will grow ever more valuable as technology automates routine work”.

Applied to translation, the development of employment trends away from routine activities towards adaptive cognitive tasks suggests that technological advances will challenge human translators where their work requires only routine cognition. The metaphorical TAUS content pyramid (Massardo et al., 2016, p. 10) depicts the current state of translation demand, dividing it into the larger segments of user-generated content, messaging, internal documentation and maintenance and support content in the broader bottom half, and user documentation, user interfaces, marketing and brand content in the narrowing top half. Much of the content in the bottom half of the pyramid is already being covered by MT, often representing new markets and large volumes that have never been or ever would be handled by human translators. Repetitive, controlled content such as user documentation and user interfaces are increasingly likely to be covered by fast-improving MT technologies, while marketing and brand content should remain the preserve of human translation. The reason is related to the growing need for quality differentiation due to greater opportunities afforded to users in choosing what languages, which content and what level of quality is important to them. The likely result for the profession as a whole is an increasing shift in the demand potential for human translators’ work towards higher-value user-centred translation, intercultural communication and transcreation (Katan, 2013, 2016; Liddicoat 2016) to enhance an organisation’s reputation, as well as ethically grounded risk management (cf. Canfora & Ottman, 2015; Pym, 2015).

That potential is most apparent at the interface between translation and corporate communications, where quality-assured reputation and trust building hinge on coherent communication strategies and concepts across an entire organisation. In order to be able to explore the synergetic opportunities more fully, a closer look is needed at the concept of strategic communication, which still serves as the guiding principle for corporate communications work in theory and practice (e.g. Argenti, 2013; Mast, 2016; Cornelissen, 2014; Holtzhausen, 2015). All organised internal and external communication activities (employee magazines, company brochures, corporate websites etc.) should be planned and designed accordingly.

Strategic communication can be defined as “communication aligned with the company’s overall strategy”, its purpose being to “enhance its strategic positioning” (Argenti, 2013). Communication obtains an important corporate function in helping an organisation achieve its strategic objectives. Strategic communication, however, has another dimension, namely the way communication is managed. Strategic management of communication ultimately serves to improve an organisation’s reputation. It ensures that the organization’s intended positive perception and its actual perception by stakeholders overlap as much as possible, thus enhancing organisational value (Mast, 2016, pp. 67-105; Quirke, 2008; Schmid & Lyczek, 2006). Derived from the concept of strategic management, strategic communication management seeks to apply a specific method to plan an organisation’s communication. This requires an initial analysis of the organisation and its environment, formulation of a communication strategy, implementation of communication measures and, finally, evaluation of those measures. Typically, the analysis, communication strategy, intended measures and evaluation methods are documented in a communications plan, providing guidance for communication practitioners by specifying communication objectives, target groups and the details of the communication strategy (e.g. core messages, preferred channels, storytelling strategy or argumentation strategy). This information is crucial in order to enable communications specialists to plan detailed communication measures, specifically by choosing appropriate communication instruments, defining the content of contributions and designing their language. Due to its reputational and value-adding impact, corporate communications has been increasingly recognised

as an important strategic function by senior management. This is why the number of communication practitioners has grown in general, and why the profession now distinguishes between communication technicians and communication managers (Cornelissen, 2014, p. 94), the latter requiring analytical, strategic and planning skills along with a solid understanding of corporate business rationale.

With increasing globalisation, companies have been confronted with the challenge of managing communication strategically not only among stakeholders in the company's country of origin, but also across all cultures in which the organisation operates. Having to cater to a variety of stakeholders with different cultural value sets, expectations and communication needs, corporate communications still has to find ways of creating a coherent international image. It needs to balance what researchers of international corporate communications call the standardisation and differentiation (Huck-Sandhu, 2015, p. 372, 2016, p. 430) of communication measures. With so much responsibility to carry, corporate communications needs to ensure that both source-text contributions and translations meet the requirements of the communication strategy in order to build an international reputation.

The authors are not aware of any substantial body of research that has been conducted on international corporate communications *per se*, although work has been done on related issues such as language contact (including English as a *lingua franca*) and multilingual discourse production (e.g. House 2015, pp. 97-115; Kranich, Becher, Höder, & House, 2011) inside and outside organisations. This, however, has not directly addressed the challenges presented to the actual processes, methods and instruments of corporate communications in international multilingual contexts, where theoretical and empirical input is scarce. Despite the observable trend among organisations to adopt *lingua franca* English as their working language, target or stakeholder groups, target markets and target cultures still need to be addressed in their local native languages in official communication. One of the main concerns articulated by practitioners is to have corporate communications material translated into other languages without being able to verify whether target texts appropriately convey the company's messages and image and thus conform to the communication strategy.

In short, as relatively new phenomena, the international and multilingual dimensions of corporate communications are only now attracting systematic attention in research and are still in the process of professionalisation. The tools and methods that communications professionals currently deploy are not necessarily adequate to handle the growing intercultural demands and complexity of strategically conceived multilingual documentation and translation products. This framework holds the very real prospect of a broader consultative and (co-)creative role for textually experienced, interculturally versed translators.

While the translation profession is in the process of repositioning itself in the face of its numerous challenges, corporate communications, we suggest, is in need of the very skills and competences that translators possess as adaptive intercultural mediators. If this potential is to be properly tapped, we need to understand the current status of the interface between translation and corporate communications. By exploring the actual and potential interplay between the two fields, we can gain new insights into how translation quality-assurance and assessment models and standards might be adapted to handle reputational texts, and how translator self-concept and education might be improved.

As things stand, leading quality-assurance models (e.g. Gouadec, 2007; Kurz, 2009) depict a linear workflow in which interlingual translation represents only one or two segments in the broader scheme of the socio-technical translation event. The linearity of these models is largely replicated in translation service quality standards such as EN 15038 (2006) and ISO 17100 (2015). In ISO 17100 (2015), for instance, translation proper is only one of twelve components encompassing pre-production, production and post-production processes. Although the models and standards have integrated feedback loops

between revisers and translators during the production phase, and from the client via project managers in the post-production phase, the feedforward from the client to the translator is restricted predominantly to technical and terminological specifications. Furthermore, while channels are envisaged for translators to ask questions via the project manager in cases of uncertainty (e.g. Kurz, 2009, pp. 151-152; ISO 17100, 2015, p. 10), there is no systematic provision for direct feedback or advice from the translator or Translation-Service Provider (TSP) to clients or authors. Kurz (2009, p. 153) does recommend that continuing professional development (CPD) and coaching measures should be put in place to allow TSPs to gain more insights into the client organisation's activities, but these he aims at the project managers rather than the translators themselves.

One of the few alternative models has been proposed by Hofmann (2012). For reasons most likely related to dissemination – the model is developed in a PhD dissertation published only in German –, it appears to have had little impact outside the German-speaking world, where it has exerted a certain influence (e.g. Canfora & Ottmann, 2015). He develops his model of translation as a product-oriented (industrial) service within the framework of business-process management. Hofmann (2012, pp. 109-110) argues that, as translation is an integral part of industrial products, the translation (business) process cannot be treated in isolation but should be integrated into an overarching information-management process encompassing knowledge management and (technical) documentation. Hofmann's detailed model (2012, p. 163-197) thus separates translation quality management out from the core translation process where other models and standards conventionally situate it. Instead, he positions it as a corporate support process, alongside terminology, content, technology, financial and personnel management, which enables it to serve integrally both the strategic and the core translation processes of a (business) organisation. Such an approach has the advantage, Hofmann (2012, pp. 181-182) states, of rendering the translation process more visible and transparent, and would allow organisations to develop more methodical, systematic process-oriented quality assurance models anchored in general industry standards. Canfora and Ottmann (2015) take up key aspects of Hofmann's approach by putting forward a comprehensive risk-management model for technical documentation and translation, based on ISO 31000. The key component of their model is a risk matrix, which enables the risk potential of source documents to be analysed and appropriate processes to be put in place to control and manage the risks and quality of the translations. Working with the results of workshops conducted with participants at two conferences, Canfora and Ottmann (2015, p. 337) tentatively propose actual measures that might be taken to lower overall risk. Most of these repeat the process stages and agencies already enshrined in standards like ISO 17100 (2015), although the perceived need to designate a point of contact for translators to directly request more information about source-document risk does appear to go beyond the mediated flow of information foreseen in the conventional standards and models.

In adopting a predominantly process-management perspective, neither Hofmann (2012) nor Canfora and Ottmann (2015) explicitly consider the deeper implications for the role of individual translators involved in these processes. However, from the viewpoint of the present article, positioning translation within corporate information management and placing the quality-management process outside the linear input-output translation process have the potential to afford a more flexible approach to quality assurance in which translation professionals might be empowered to broaden their role. Furthermore, Canfora & Ottmann (2015, p. 340) do point out that applying a risk-management model to translation could well enhance the reputation of the translation profession by increasing clients' attention towards the key significance of translation's risk potential.

Notwithstanding the fresh and promising approaches that Hofmann and Canfora & Ottmann have embarked upon, there does seem to be a blind spot in the currently prevailing models and standards to the active value-adding role that translators themselves can play in providing feedback and advice to source-text authors, clients and other actors in the socio-technical process of translation. This is also reflected in current approaches to quality assessment. The influential integrative quality-assessment model proposed by House (2015), for example, tends to cement the view of translators bringing their expertise to bear almost exclusively at the juncture between themselves and their texts. House uses the example of a genre – a mission statement embedded in an annual report – that is crucial in representing a company’s identity and justifying its existence towards society. Yet, the “statement of quality” of the translation is based on source-text, target-text and parallel corpus analysis (House, 2015, pp. 129-141) alone, wholly neglecting the client organisations’ corporate communication strategy.

The role that such models and standards assign to translators reflects a profession characterised by systemic constraints and restricted responsibilities, and the situation will only worsen if the role and position of human translation are inadequately addressed by translator education. The understandable educational response to technological change in general, and MT in particular, has been to introduce the latest technologies across many parts, if not all, of the curriculum. Yet, an uncritical focus on technology seems likely to entrench the risks of fixation, stagnation and educating low-autonomy pre-professionals, and may serve to distract educators from the value-adding potential that distinguishes the adaptive expertise of human translation from routine activity. As Katan (2016) succinctly puts it, “technical, low-risk, low ambiguity translating and interpreting can be safely delivered with minimum human intervention” (p. 377), which Canfora and Ottmann’s model (2015) also suggests, but it is only translators that are professionally trained as intercultural mediators and communicators who are and will be “uniquely placed to intervene and mediate between contexts to ensure optimum communication” (p. 337). If the profession is to be equipped for the future, it will have to adapt – and with it the institutions, curricula and teachers that educate for it.

RESEARCHING FOR THE FUTURE: INVESTIGATING QUALITY-RELEVANT PROFESSIONAL PRACTICES AND SELF-CONCEPT

Motivation: Pre-Professional Responses and Practices

Recognising this need, the MA in Applied Linguistics at Zurich University of Applied Sciences, with its unique combination of specialisations in Professional Translation, Conference Interpreting and Organisational Communication¹, has decided to develop large-scale real-life collaborative projects combining the three specialisations in the design and application of corporate communications concepts. The projects were introduced to an existing module, “Applied Linguistics II”, weighted with 6 ECTS credits and attended by students generally in their second semester of a three-semester programme. As such, the translation students involved qualify for the increasingly accepted designation “pre-professional” (Hagemann, 2016, p. 37).

Under the guidance of teachers from all three disciplines, students of each specialisation work together to develop a public event derived from an overall communication strategy. The students’ task is to stage a panel discussion with public participation, bringing together multilingual communications experts from university departments and industry. The overriding purpose has been to sensitise the students to

the exigencies of their professional core tasks, break down traditional task boundaries between them and, by pooling cross-disciplinary competences, develop an extended role perception, or self-concept, amongst the three student groups.

Student quality-assurance evaluations and feedback sessions were conducted on two course iterations in the final quarters of 2015 and 2016. These comprised an anonymous online questionnaire survey, the standard instrument used to evaluate courses on our programmes, which combines closed questions requiring responses on a four-point Likert scale (containing ratings from A, coded as 1.00, the highest quality assessment category, to D, coded as 4.00, the lowest) with the opportunity for participants to add voluntary comments. A total of 55 (out of 92) students took part in the surveys, 29 in the first and 26 in the second survey, representing a response rate of 58% and 62% respectively. The first survey was also followed by a semi-structured group interview with one representative of each of the three specialisations. Two teachers moderated this 90-minute session, which took place approximately one month after the event, and recorded responses in written summary form. Students were encouraged to offer comments on the project and on aspects of the evaluation results, with a particular emphasis on collaboration across the three specialisations, though no strict order of questions and answers was observed.

The quantitative responses from the survey of the 2015 course provided a mean score across the questions for the course, including the project, of 1.93, which stood just above the value for the B grade (2.00). This was rather higher than that of the previous student evaluation of the course (n=25), which had no project, when the mean was 2.38. The value of the practical project component was repeatedly stressed in the comments made by the students on the 2015 course. The quantitative responses from the survey of the 2016 course showed a slightly lower evaluation score of 2.08, marginally below the value for the B grade (2.00).

The qualitative data from the open responses and the group interview repeatedly stressed the value of the practical project, and they also indicated that the students from all three specialisations, when made aware of the guiding principles of strategic communication, clearly acknowledged their importance for the work they were doing. However, it was equally apparent that such awareness first needed to be planted through the teachers' interventions and questions, both during the coaching sessions and in the focus-group interview. In fact, the comments identifiably made by the translation and interpreting students clearly demonstrated that they remained narrowly focused on their traditional core fields of activity, and observations by the teachers involved in the project strongly support this impression. Presumably, the students had been unable to identify ways of integrating specific knowledge of strategic communication management into their processes, which seems to be related to inadequate signalling of learning objectives and over-optimistic expectations placed on learner autonomy, especially in the second iteration. It will therefore be necessary for the teachers to step up their scaffolding and signposting techniques in future.

Moving beyond the immediate pedagogical utility of this pre-professional survey and interview data, we began to conjecture that the results could well reflect the situation in the professions themselves. This view was sustained by our enduring contact with the professional fields of translation and corporate communications and backed further by observations and informal enquiries among freelance and institutional translators, TSPs and corporate communications units.

A pilot study was therefore launched to investigate, firstly, the interactions between translators, translation project managers and communications specialists; and secondly, whether translators' current self-concepts give them the potential to adopt a broader consultative and (co-)creative role within the framework of strategic communications, with corresponding implications for education and professional development. The study, initially restricted to participants working in Switzerland, is intended to form

the basis of an international follow-up project aimed at combining models from translation and corporate communications with the ultimate goal of adding new qualitative and theoretical dimensions to both fields.

The researchers' working hypotheses were as follows:

1. There is only limited awareness amongst translators and project managers of the communication strategies and/or communications plan of the clients and/or organisations requesting their services.
2. Communications specialists have only limited awareness of the role translation can play in their organisation's communication strategy and/or the implementation of its communications plan, reflected in the low levels of relevant feedback they give to translators and/or project managers.
3. Only limited feedback is given to communications specialists by translators and/or project managers on the strategic adequacy of source texts.

Study Design

Data for the study was gathered in two ways. Firstly, an anonymous online survey was launched in June 2017² to collect quantitative and qualitative data from the following participant groups working in Switzerland: freelance translators, institutional staff translators, commercial staff translators, institutional translation project managers, commercial translation project managers and communications specialists working in the public and private sectors. The authors decided to conduct two separate surveys because they wished to examine the interfaces between the translation profession and corporate communications from the different perspectives of the two professional groups. Two basic questionnaires were therefore used: one for translators and project managers, made available in four languages (English, French, German and Italian), the other for communications specialists, supplied in two languages (English and German). The call for translators and translation project managers to participate was distributed within Switzerland through known contacts in industry and public administration, lists of professional participants in the CPD courses offered at Zurich University of Applied Sciences, alumni organisations and professional associations for translators, interpreters and terminologists. The call that went out to the communications specialists was distributed through a mailing list containing the contact addresses of corporate communications units at the top 100 Swiss companies³, on the one hand, and at approximately 100 leading Swiss public institutions (public administrative offices, local, cantonal and central government agencies, universities, etc.), on the other. Distribution of all calls took place by e-mail, which contained only cursory information explaining the survey's purpose so as to minimise the risk of introducing a response bias. Data was also collected in three in-depth semi-structured interviews with one representative of each of the three translator work profiles. Due to space restrictions, the interview data will be presented in a future publication.

The questionnaires in the online surveys largely contained identical questions with wordings adapted to accommodate the different target groups, but some additional questions were also directed solely at one professional group or the other. The questionnaires comprised optional questions to elicit personal information and mandatory questions on professional experience, work profile, awareness of and access to communication strategies and communications plans, as well as questions addressing the interaction of translators and corporate communications specialists (see Appendices 1 and 2 for the questionnaires used). All questions apart from those on professional experience and work profile were answered on a four-point Likert scale with the possibility of selecting the additional "not applicable" option. The response scale was either "fully", "mostly", "partly", "not at all" or "I agree fully", "I agree mostly",

“I agree partly”, “I do not agree”. In some cases, the respondents were asked to specify or expand on their choices. All participants except the freelance translators were asked questions about the industry/sector and regions in which the organisations they worked for operated. All questions were formulated in accordance with the best practices for questionnaire surveys recommended by Rasinger (2013, pp. 59-85) and Saldanha and O’Brien (2014, pp. 151-158). The relative novelty of the present pilot study and survey precluded the use of questions from pre-existing (validated) scales.

Finally, all translator categories were asked how they saw their principal role(s). The response options and scale were based loosely on those developed by Göpferich (2008, pp. 266-267) and adapted to accommodate the cline of professional self-concept categories proposed by Ehrensberger-Dow and Massey (2013), itself derived in part from a continuum sketched out by Kiraly (1997, p. 152). The categories were extended to include the concepts of mediation, co-creation and providing feedback on source texts, which the authors see as key to the extended role that translators might assume. The precise wording of the statement that the respondents had to complete was “As a translator, I see my principal role(s) as”. Respondents also had the option of adding a description of how they viewed their role, either to supplement previous choices or as an alternative to the options offered.

Results

Translators and Translation Project Managers

At the end of July 2017, 226 respondents had completed all or part of the on-going survey circulated among translators and translation project managers. As approximately 34% failed to complete the questionnaire fully, a varying number of participants responded to each question. In this section, we shall therefore summarise the data collected from those respondents who completed at least the first two survey questions (n=190). The results for the responses on the Likert scales will be presented as percentages of the total number of answers to each question, which does not include the responses of participants selecting the “not applicable” option. To maximise transparency, the questions and the number of applicable responses to each are listed in Table 1. The table does not contain the survey question on the translators’ views of their roles, since participants could give multiple responses to this question.

Among those participating in the translators’ and project managers’ survey, awareness of the communication strategy of the clients and/or organisations they worked for appears to be high. Sixty-three per cent stated that they were either fully (17%) or mostly (46%) aware of it, and 30% partly aware. Only 7% were not at all aware of the communication strategy. However, access to the communications plan – in other words to the written form of the strategy – seems to be more restricted, with the comparatively lower figure of 39% saying that they agreed fully (12%) or mostly (27%) that they had access to the plan or plans, whereas 35% partly agreed and 26% did not agree.

On the issue of the advice they received on how they might contribute to clients’ or employers’ communication objectives, 35% agreed fully (9%) or mostly (26%) that they received general advice, while 65% agreed only partly (37%) or did not agree (28%). Specific linguistic advice appears to be even sparser, with 30% agreeing fully (8%) or mostly (22%) that they received it, and 69% agreeing only partly (47%) or not at all (22%).

In respect of the qualitative data on the types of advice received, it is clear that conventional style guides are by far the most common form of feedforward (51 mentions), followed by terminology databases, glossaries and reference documents such as parallel texts (44 mentions). Less frequently, informa-

Quality Assurance in Translation and Corporate Communications

Table 1. Translators and translation project managers: Number of applicable responses per survey question

Question	Total Responses
Years of professional experience to the nearest year	190
Please select the statement below that most closely describes your work profile: I am a translator who mainly works freelance / as a staff translator for government/public institution or NGO / as a staff translator for a commercial organisation. I am a translation project manager who works for a government/public institution or NGO / for a commercial organisation.	190
I am aware of the communication strategy of the client/organisation(s) I work for.	174
I have access to the communications plan (i.e. the written form of the communication strategy) of the client/organisation(s) I work for.	165
I receive general advice on how to contribute to the communications objectives of the client/organisation(s) I work for.	163
- Please specify in brief the sort of advice you have received.	75
I receive specific linguistic advice on how to contribute to the communications objectives of the client/organisation(s) I work for.	156
- Please specify in brief the sort of advice you have received.	76
I can ask the client/organisation(s) directly any questions I have about the communication strategy.	151
I can ask translation project managers any questions I have about the communication strategy.	103
I receive feedback on the strategic adequacy of our/my texts from the client/organisation(s).	148
I receive feedback on the strategic adequacy of our/my texts from source-text writers.	146
(Only for freelance and staff translators) I receive feedback on the strategic adequacy of our/my texts from translation project managers.	102
(Only for translation project managers only) I receive feedback on the strategic adequacy of our/my texts from translators.	25
(Only for staff translators and translation project managers) I work for an organisation that operates mainly in Switzerland / internationally.	73
(Only for staff translators and translation project managers) I work for an organisation in the following industry/sector: Manufacturing industry / Trade / Financial services, insurance / Energy / Building, housing / Tourism, hospitality / Healthcare / Education, research / Transportation / Media / Information, communication / Public administration / Other ...	71

Source: (Survey of Translators, Translation Project Managers and Corporate Communications Specialists 2017).

tion on the target audience, the expected audience design and the degree of fidelity to the source text is communicated (18 mentions in all). Information on the context, purpose and layout of the target text (8 mentions) as well as direct instructions on tonality, voice and register (8 mentions) appear scarcer, as do the more formal, mechanical aspects of corporate identity such as corporate wordings, and corporate communications or design manuals (10 mentions). More pertinently, access to communications objectives, communications plans or a corporate philosophy (5 mentions) as well as input from, or collaboration with, corporate communications on communication strategies (4 mentions) are only rarely indicated.

The culture and channels for asking questions about the communication strategy present a rather different picture, with 64% of translators and project managers agreeing fully (35%) or mostly (29%) that they can ask the client or employer directly. Twenty-five per cent partly agreed and only 11% did not. Sixty-seven per cent of translators agreed fully (36%) or mostly (31%) that they can ask their project managers, with 20% agreeing partly and 13% not agreeing.

When it comes to receiving feedback on the strategic adequacy of texts, the results indicate room for improvement. Only 26% agree fully (6%) or mostly (20%) that they receive it from clients and employers,

only 15% agree fully (3%) or mostly (12%) that they get it from source-text writers and only 17% agree fully (1%) or mostly (16%) that they receive it from project managers. 74% agree only partly (38%) or not at all (36%) that they get feedback from clients or employers, 85% disagree (51%) or only partly agree (34%) that they receive it from source-text authors, and 83% agree partly (40%) or do not agree (43%) that they obtain feedback on the strategic adequacy of their work from project managers.

The results for the independent variables in our survey – experience, work profile and activity sector – are mixed. When grouped into five-year intervals, the participants' experience range is fairly evenly spread across the intervals between one and 25 years, gradually tailing off thereafter: 18% indicated 1-5 years' professional experience, 18% 6-10 years, 13% 11-15 years, 19% 16-20 years, 15% 21-25 years, 8% 26-30 years and the remaining 8% 31 years or more. The respondents' declared work profiles reveal that freelance translators account for the majority of participants, at 48%, commercial staff translators and commercial translation project managers make up 21% each, 9% are institutional staff translators and a mere 1% (n=2) represent institutional project managers. Finally, the sectors listed by the staff translators and project managers show that approximately 53% worked for organisations operating in financial services (39%), manufacturing (7%) and trade (7%), 14% in public administration, 10% in information and communication (7%) or the media (3%), 4% in education and research, and 3% in healthcare. The remaining respondents stated that they worked either for translation companies or in specialised sectors such as security, housing or hospitality.

What relationships can be determined between these independent variables and the dependent variables awareness, feedforward, culture and channels for asking questions, and feedback? While no discernible pattern emerges in relation to experience level, the relationship between these variables and work profile does yield some noteworthy results. A chi-squared test of independence indicated statistically significant differences ($p < 0.01$) for three sets of responses: on the awareness of communication strategies, on access to the communications plan and on feedback from source-text writers about the strategic adequacy of translations.

Only 50% of the freelance translators declared themselves fully (9%) or mostly (41%) aware of their clients' communication strategies, with a far higher proportion shown for staff translators (87%; 20% fully and 67% mostly) and project managers (100% fully) working for government/public institutions or NGOs, as well as for commercial staff translators (79%; 32% fully and 47% mostly) and commercial project managers (66%; 16% fully and 50% mostly). Unsurprisingly, 73% of freelancers agreed only partly (36.5%) or not at all (36.5%) that they had access to client organisations' communications plans. This compares with 51% of commercial translators who claimed they had only partial (32%) or no (19%) access, and 68% of commercial project managers (49% agreeing partly, 19% not at all). By contrast, the institutional translators and translation project managers appear to have far more access, with 80% of translators agreeing fully (13%) or mostly (67%) and one project manager each agreeing fully (50%) or mostly (50%).

When asked whether general advice on contributing to organisations' communications objectives was fed forward to them, 50% of the commercial staff translators agreed fully (19%) or mostly (31%) and 39% of commercial project managers did the same (15% fully, 24% mostly). The other three groups appeared to receive much less advice: 71% of freelancers agreed only partly (44%) or not at all (27%), 66% of institutional translators (33% partly, 33% not at all) and both institutional project managers (100% partly) doing the same. Although the latter did claim that they mostly (100%) received specific linguistic advice (with none agreeing fully, partly or not at all), the rest of the respondent groups produced less positive results for this question, led by the commercial project managers (50% agreeing only partly

Quality Assurance in Translation and Corporate Communications

and 27% not at all) and the freelancer translators (50% agreeing only partly and 22% not at all), and followed by the institutional translators (29% agreeing partly and 36% not at all) and the commercial staff translators (49% agreeing partly and 11% not at all).

The issues of asking questions and receiving feedback presented a conflicting picture. All work-profile groups overwhelmingly agreed fully or mostly that they could directly ask questions about the communication strategies of the clients or organisations they worked for. Likewise, a clear majority of freelance and commercial staff translators also agreed fully or mostly that they could ask the same of their project managers (though only 50% of institutional translators shared this opinion). By contrast, the vast majority of all groups, apart from the two participating institutional project managers, responded that they received only partial or no feedback on strategic adequacy from their clients or employer organisations, the source-text writers or, in the case of the translators, from their project managers. The same applies to feedback *from* translators, with 86% of the project managers not agreeing (43%) or agreeing only partly (43%) that they received it. It seems that, whereas the possibility and preparedness to ask questions is most definitely given, there is no equivalent feedback culture, at least not at the strategic level of textual adequacy.

If we now examine the results for the small number of respondents working in the public administration sector (n=10) in Switzerland, a slightly different pattern emerges. As many as 89% of this group agreed fully (33%) or mostly (56%) that they are aware of their organisations' communication strategy and 67% fully (11%) or mostly (56%) claimed access to the communications plan – as opposed to 63% and 39% for the overall participant sample. Yet, 90% responded that no or only partial general advice on how to contribute to communications objectives was fed forward to them, compared with 65% of the overall sample, while the figure for receiving specific linguistic advice remained the same as for the overall sample, where 70% agreed only partly or not at all.

A higher proportion (80%) than in the overall sample (64%) agreed fully (60%) or mostly (20%) that they could ask their organisations directly about the communication strategy, although only 25% of those to whom the question applied (n=4) agreed mostly, and none fully, that they could ask their project managers (the figure was 67% in the overall sample). Regarding feedback on the strategic adequacy of their work, the results are in keeping with those for the overall sample: none agreed fully and a mere 20-22% agreed partly that they received it from clients, employers and/or source-text writers, and none agreed either fully or partly that they received it from their project managers. Only one project manager responded to the question about feedback from translators on the strategic adequacy of the source texts, stating that he agreed mostly that he received it. So it seems that, despite a higher awareness of the communication strategy than the rest of the sample shown by those working in the public sector, advice fed forward was rather lower, and feedback equally low.

Finally, we turn to the results for the self-concept variable. The participating freelance and staff translators were asked to indicate which of the ten choices offered reflected how they saw their principal role or roles, with multiple responses possible. The precise wording of the statement they had to complete was “As a translator, I see my principal role(s) as ...”. Respondents also had the option of adding a description of how they viewed their role, either to supplement previous choices or as an alternative to the options offered. The results for the closed questions are presented in Table 2 in descending order of the total number and percentage of responses for each option. Since 5 of the original 9 open responses very closely matched some of the closed options, the figures have been adjusted accordingly.

The remaining individual open answers were: verifying the correctness of texts published in the target language, battling with “Globish”, liaising between the technical office and the worksite and, finally,

Table 2. Number and percentage of responses on translators' roles

Question As a translator, I see my principal role(s) as ...	Total Responses per Option (n=782)	% Responses per Option
remaining faithful to the source-text writers' intentions.	115	15
fulfilling my clients' wishes and/or requirements.	105	13
meeting document specifications and/or demands.	94	12
transferring the main messages of the source text.	92	12
meeting project management standards and/or deadlines.	89	11
helping clients communicate messages and/or intentions.	87	11
transforming source texts for my target audiences.	84	11
mediating between language and cultural groups.	56	7
providing feedback on source texts and/or assignments.	32	4
co-creating texts with clients and/or source-text authors.	24	3
Other	4	1

Source: (Survey of Translators, Translation Project Managers and Corporate Communications Specialists 2017).

producing effective English-language corporate communication by positioning translators at the centre of company communications⁴.

Overall, then, overt mediatory, co-creative and advisory self-concept ranked lowest, whereas the less autonomous categories of fidelity to source-text writers' intentions and meeting client requirements, document specifications and project-management standards scored highest. Nevertheless, virtually equal weight is given to helping clients communicate their messages and intentions and to transforming source texts for target audiences, suggesting the strong general presence of both an assistive and an adaptive role awareness.

When these results are related to the independent variables, the experience data again reveal no obvious pattern. However, work profile does seem to have a bearing on the results. A higher proportion of freelancers (82%) than institutional (72%) or commercial (73%) staff translators indicated that fidelity to source-text writers' intentions constituted a principal role. But a lower proportion of freelancers saw (adaptively) transforming source texts for target audiences as a main role, 47% as opposed to 63% of commercial and as many as 82% of institutional staff translators. The freelancers also scored slightly lower on co-creation, at 13% compared with 18% of institutional staff and 23% of commercial staff translators – though, as we have already seen, the result for all three groups is comparatively low in the first place. The institutional staff translators scored lowest on meeting document specifications and demands (47%, compared to 62% of freelancers and 73% of commercial staff translators) and on fulfilling clients' wishes and requirements (53%, as opposed to 71% of freelance and 73% of commercial staff translators), and highest on the mediatory role between languages and cultures (53%, compared with 36% of the freelancers and 38% of the commercial staff translators). At 47%, fewer institutional staff than freelance and commercial staff translators (60% each) also saw themselves in the assistive role of helping clients communicate their messages and intentions. Finally, while the overall figure for all three groups is admittedly low, it is the commercial staff translators who show the lowest proportion of those providing feedback on source texts and assignments, at 13% (compared to 24% of freelance and 29% of institutional staff translators).

Analysing the relationship between the independent work-profile variable and the self-concept variable, we can say that the freelancers appear to see their role less in transforming and adapting texts and (slightly) more in remaining faithful to the source text than the other groups. Indeed, it is the institutional staff translators who appear to see themselves in the more autonomous roles of text adaptors and linguistic/cultural mediators, and they also seem to be less bound to document-specific and client-specific demands and requirements than the other groups. About two fifths of them also define providing feedback on source texts as part of their self-concept, a low figure, but one which compares favourably with those for the other groups. However, the overt client orientation of the institutional translators is less in evidence than amongst their freelance and commercial staff counterparts, as they appear not to give the same priority to assisting clients with their messages and intentions. This is perhaps not surprising, given the direct economic necessity of maintaining good customer relations imposed on freelancers and commercial staff translators alike.

Communications Specialists

The survey of communications specialists was fully (40) or partly (33) completed by 73 respondents, who consisted of two types of professionals – those who frequently request translations and those who write texts for external and/or internal PR. As in the previous section, we will present the results for the responses on the Likert scales as percentages of the total number of responses to each question, excluding the “not applicable” ones and including only those respondents who completed at least the first two survey questions (n=59), with the total number of responses to each question listed in Table 3. Of the 59 respondents, 25 indicated that they frequently request translations and 34 stated that they write PR texts.

As far as the survey among all corporate communications professionals is concerned, awareness of their employer’s communication strategy is, as expected, very high. Ninety-three per cent stated that they were either fully (58%) or mostly (35%) aware of it. Only 7% were merely partly aware of the communication strategy. No respondents indicated that they were not at all aware. 95% of all communications professionals stated they had access to the communications plan (78% agreed fully, 17% mostly), while 5% agreed only partly (3%) or did not agree (2%). Thirty-five per cent claimed that they gave advice to TSPs on how to contribute to the communications objectives, with 21% agreeing fully and 14% mostly. This means that 65% of all communications professionals questioned do not instruct translators on the corporate communications objectives (42%) or only do so to a certain extent (23%). When asked if they gave linguistic advice to TSPs on how to contribute to the communications objectives, 46% agreed fully (16%) or mostly (30%). 54% agreed only partly (16%) or did not agree at all (38%).

However, looking at the type of general advice corporate communications professionals give to TSPs regarding communications objectives, we see that this in fact consists mainly of normative linguistic advice in the form of style guides, terminology databases, corporate wordings and so on. Only 5 respondents named advice that indicated an understanding of the text as an integral part of the organisation’s strategic communication, namely the “objective and purpose of the translation”, “the text’s relation to the communication strategy”, the “context of the translation and purpose of the text”, the “target group of publication, channel and existing documents” and “annual targets and measures”. When the respondents were asked to specify the linguistic advice they provided to translators, their responses largely included the same type of advice as listed in the question on general advice: terminology, spelling conventions, style guides, wordlists, tonality instructions and so forth.

Table 3. Corporate communications professionals: Number of applicable responses per survey question

Question	Total Responses
Years of professional experience to the nearest year	59
Please select the statement below that most closely describes your work profile: I work as a corporate communications professional who frequently requests translations / who writes texts for external and/ or internal PR.	59
I am aware of the communication strategy of the organisation I work for.	58
I have access to the communications plan (i.e. the written form of the communication strategy) of the organisation I work for.	54
(Only for corporate communications professionals who write PR texts) I receive general advice on how to contribute to the communications objectives of the organisation I work for.	28
- Please specify in brief the sort of advice you have received.	18
I give general advice to translation service providers on how to contribute to the communications objectives of the organisation I work for.	43
- Please specify in brief the sort of advice you have received.	15
(Only for corporate communications professionals who write PR texts) I receive specific linguistic advice on how to contribute to the communications objectives of the organisation I work for.	19
- Please specify in brief the sort of advice you have received.	9
I give specific linguistic advice to translation service providers on how to contribute to the communications objectives of the organisation I work for.	37
- Please specify in brief the sort of advice you have received.	16
(Only for corporate communications professionals who write PR texts) I can ask the organisation directly any questions I have about the communication strategy.	18
(Only for corporate communications professionals who write PR texts) I receive feedback on the strategic adequacy of our/ my texts from the organisation.	16
I receive feedback on the strategic adequacy of our/my texts from translation project managers.	29
I receive feedback on the strategic adequacy of our/my texts from translators.	33
I work for an organisation that operates mainly in Switzerland / internationally.	40
I work for an organisation in the following industry/sector: Manufacturing industry / Trade / Financial services, insurance / Energy / Building, housing / Tourism, hospitality / Healthcare / Education, research / Transportation / Media / Information, communication / Public administration / Other ...	36

Source: (Survey of Translators, Translation Project Managers and Corporate Communications Specialists 2017).

When asked whether they received feedback from translation project managers on the strategic adequacy of their texts, only 7% of all communications specialists either agreed fully (3.5%) or mostly (3.5%); 93% agreed partly (21%) or did not agree (72%). Translators seem to give even less feedback to communications professionals than do translation project managers: 0% of all communications professionals agreed fully and only 6% mostly, while 94% agreed partly (15%) or did not agree (79%).

While the results generated across both work profiles are not really surprising, an analysis of the data by work profile reveals some unanticipated, yet plausible, insights. Forty-one per cent of corporate communications professionals commissioning translations agreed fully (27%) or mostly (14%) when asked if they gave general advice to TSPs on how to contribute to the communications objectives. Fifty-nine per cent agreed only partly (32%) or did not agree (27%). On the same question, only 28% of those who write PR texts agreed fully (14%) or mostly (14%), with 71% agreeing only partly (14%) or not agreeing (57%).

The question of whether they give specific linguistic advice to TSPs on how to contribute to the communications objectives was also answered rather differently by the two work profiles. While 57% of communications professionals who request translations agreed fully (24%) or mostly (33%), only 31% of copywriters agreed fully (6%) or mostly (25%). Accordingly, 43% of communications professionals who request translations agreed only partly (14%) or did not agree (29%) whereas 68% of all copywriters agreed only partly (18%) or did not agree (50%).

When it comes to receiving feedback from translation project managers about the strategic adequacy of their texts, the results of the two work profiles do not differ very much. Among those who request translations, none agreed fully, 6% mostly, 23% partly and 71% did not agree at all. In comparison, 8% of the copywriters agreed fully, none mostly, 17% partly and 75% did not agree at all.

The issue of feedback from translators about the strategic adequacy of their texts was also assessed very similarly by both work profiles: none of the communications specialists who request translations agreed fully, 5% mostly, which means that 95% agreed only partly (16%) or did not agree (79%). By comparison, no writers of PR texts agree fully and 7% mostly, while 93% agreed partly (14%) or did not agree (79%).

DISCUSSION AND IMPLICATIONS: RE-CONFIGURING SELF-CONCEPT AND RE-EDUCATING FOR THE FUTURE

Discussion of Survey Results

The overall results of the translators' and project managers' survey indicate a degree of awareness of communication strategies that does not support the corresponding part of the first hypothesis. On the other hand, a much lower proportion of respondents had access to the communications plan. As the communications plan is the written form of the communication strategy, the result is somewhat puzzling, and leaves room for interpretation. One possible explanation is that the respondents have a vague idea of how the company wants to communicate but might not know what sort of detailed information a proper communication strategy contains, suggesting that more could be done to instruct translation professionals in this area. Another possibility is that no formal communications plan exists in the first place, though in this case we would have expected respondents to have chosen the "not applicable" option (which some of course did). At all events, this rather odd discrepancy warrants further investigation in the planned follow-up study.

The second hypothesis seems to be clearly supported by the results from the translators and project managers, in that only a third agreed fully or mostly that they received general advice, and even fewer that they received specific linguistic advice. Moreover, translators and project managers appear to receive little feedback on the strategic adequacy of their target texts. These findings are matched by the results of the survey of communications specialists on general advice, though just under half of these groups of respondents did agree fully or mostly that they passed on specific linguistic advice to the translators and project managers they had commissioned. Yet, a closer look at the nature of the linguistic advice passed on by the communications professionals, and the advice received by the translators and project managers, shows it to consist in large part of normative instruments such as style guides, corporate wordings, terminology and the like. This is not sufficient for a translator to be able to contribute to implementing a communication strategy, serving simply to help maintain a corporate identity. Information on an ar-

gumentation strategy, for example, would enable translators to make a real contribution to communications objectives. But lacking the existence of appropriate feedforward processes, it is probably easier to pass on style guides and corporate wording instructions than to explain a communication strategy and its implementation. Furthermore, communications specialists who request translations appear to give more advice, both general and linguistic, than the writers themselves do. Although this probably has to do with established work processes and routines that make requesters the single point of contact with TSPs, it would certainly make much more sense if those who actually wrote the text assumed this task, given that, ideally, they have already had to consider the strategic objectives during their writing process. In all the above cases, awareness could easily be increased and productive feedback loops fostered by designing and implementing specifically tailored processes.

With respect to our third and final hypothesis, it appears that translators, in particular, do not provide substantial feedback on the strategic adequacy of their source texts. The survey of communications specialists shows that there is hardly any feedback from translation project managers and translators, while translation project managers themselves report a similarly low level of feedback from translators. This may well be related to our previous assumption that translators (and project managers) generally have only a vague notion of what actually constitutes a communication strategy and communications plan. It is also likely, of course, that these low figures are at least in part related to the fact that feedback from translators and project managers is not part of the official work process, as we have seen in relation to the quality standards and models touched on in the Background section. This can and should most obviously be remedied by creating direct channels for structured feedback from translators to client organisations and source-text authors. After all, translators are ideally placed to play a significant part in international corporate communications, given that corporate communications specialists are usually unable to assess the quality of translations that largely determine an organisation's image abroad.

However, an equally plausible interpretation can be derived from our findings on the translators' self-concept. As already mentioned above, the overall data on how the translators see their roles demonstrate that, despite a fundamentally assistive and adaptive role awareness, mediation, co-creation and advising do not feature large in the way the translators regard their role. This tendency, which is especially – though by no means exclusively – evident among the freelancers involved in the survey, holds a number of implications, particularly in view of the predicted shifts in the human translation market described in the Background section.

Implications for Quality and Translator Education

The previous sections have not only demonstrated how important it is for corporate text production to adhere to and integrate the mechanisms of strategic communication management, they have also repeatedly drawn attention to the untapped potential for better collaboration between translators and communications specialists. We would therefore like to suggest three sets of measures to improve translation quality assurance, and with it professional opportunities for translators, by facilitating interactions between corporate communications and translation.

The first of these hinges on knowledge of the communication strategy. In order to fully understand the motivation behind source texts, translators need to be familiarised with the concept and key elements of a communication strategy. They should also have consistent and sustained access to the organisation's communications plan, with information on its core messages, target audience, and communication objectives, and receive structured, consistent and detailed advice on how to achieve the communications

objectives by linguistic means (storytelling strategy, argumentation strategy, tonality, corporate wording and so on). Unfortunately, few organisations appear to have this information readily available, as it is not especially common to formulate a communications plan in such detail. To remedy this shortcoming, we propose adopting the method of “message design” (Perrin & Stücheli-Herlach, 2013: 25-34), which provides a framework to help practitioners formulate corporate communications texts based on the organisation’s core messages. Finally, processes should be designed that grant translators systematic access to communications specialists who understand and can explain the communication strategy. These specialists need to have up-to-date knowledge of the on-going strategy process and other organisational developments that have a direct impact on the communication strategy. To the authors’ best knowledge, no such provisions currently exist in key translation quality standards and quality-assurance models.

The second set of measures has already been covered in the previous section and focuses on designing and implementing direct, enhanced feedforward and feedback processes between translators, communications specialists and source-text authors. At this point, it is important to stress that it is not only translators who can benefit from the hidden potential at this interface. Those having to formulate corporate communications texts are often not the same as those responsible for the communication strategy or involved in the strategy-making process, and they are rarely given the organisation’s communications plan during their briefing. They must often write the source texts without essential information for corporate text production, drawing inspiration from other accessible corporate material like websites or brochures. These shortcomings are often only revealed when translators come into play and work on source texts. Assuming that corporate communications will necessarily have to implement quality assurance measures during source-text production, it is feasible that translators could assume the role of quality assurers, providing feedback to corporate communications managers about strategy-related source-text inadequacies.

The third and last set of measures addresses new roles and professional opportunities for translators. With translation acquiring an increasingly strategic function in organisations (cf. Massardo et al., 2016, p. 10), one palpable implication is that it should be seamlessly integrated into strategic communication management and plans. A related proposal is touched on by Hofmann (2012, p. 205). In considering the implications of his business-process model for translation services (see above), he suggests that future research could be directed towards enlarging the theoretical underpinnings of TPR, hitherto focussed mainly on internal cognitive processes, to embrace strategic aspects of translation process management. He points out that the functional, cognition-oriented concept of translation strategies first described by researchers like Hönig and Kussmaul (1982) might be beneficially revised to accommodate factors governing strategic corporate decision-making. Though Hofmann does not expand further, the third set of measures implied by the present study can be viewed in the similar light of positioning translation and translation professionals closer to the strategic level of corporate communications

Corporate communications specialists are able to assess the strategic quality of the source texts, but are usually unable to assess the quality of translations. Yet those are the texts that largely determine an organisation’s image abroad. As textually experienced receivers and producers of texts, translators themselves should be well placed to provide informed feedback on the quality of their source material, and to advise on multilingual texts intended for international consumption. The intercultural mediation inherent in their work (Liddicoat, 2016) affords them an ideal opportunity to play an integral part in assessing and assuring output quality in internationally oriented, multilingual corporate communications. The very fact that a number of TSPs have now added editing and writing services to their portfolio⁵ can be seen in the context of the broader role that translators and related language workers are potentially

able to adopt. Indeed, if translators were included in the text creation process at an early stage, they could provide information about cultural concepts and stakeholders' needs in the organisation's markets, and might even assume a co-creative role. In helping to realise the full potential of corporate communications, translators could re-define and broaden their professional opportunities and range, developing an extended self-concept as intercultural mediators, adaptive transcreators and language consultants. The time and effort involved in providing such value-added services would, of course, require appropriate remuneration. We anticipate that, if translators and human translation succeed in repositioning themselves in this way, it will go a long way towards correcting the recent erosion of professional status (European Commission, 2012) noted above and should have a concomitant effect on compensation.

Currently, international communication is usually conducted by means of designing a communication strategy in the company's home market and adapting it to other countries and cultures. Corporate communications managers pass on the relevant strategic information to their subsidiaries or communication agencies in the target markets and supervise them as they derive appropriate measures or, in some cases, first make cultural adaptations to the strategy. Translators are employed to translate not only the strategic texts for the agencies abroad, but also those texts intended for the stakeholders' consumption that do not need any content-related cultural adaptation. However, as the world becomes increasingly globalised, corporate communications will progressively have to develop towards a truly multicultural discipline. This would also require culturally experienced, multilingual communications experts who are able to design corporate communications internationally and multiculturally right from the beginning. Such demand for international communications specialists could open up entirely new professional perspectives for translators, who would have to acquire some corporate communications skills to supplement their core competencies.

Clearly, then, these new roles and opportunities emerging for translators have significant consequences for how translator education should prepare students and practitioners engaged in continuing education for the work markets of the future. The essentially situated activity of translation makes it wholly appropriate to base curricula on authentic experiential learning (cf. Kiraly, R  th, Canfora, & Cnyrim, 2016). Translator education institutes should therefore continue to expose their students to the technological developments that will necessarily shape and support their future professional lives. But students should also be consistently encouraged to reflect on the potential and the limitations of deploying those technologies and, above all, be made aware of the added value that human translation can hold.

They should also be equipped to deliver that value. Translator education must resist the current trend towards an increasingly dominant focus on assistive technologies and foster an extended self-concept by developing students' intuition, creativity, ethical judgement and advisory skills. Greater weight should therefore be given to transcreation, user-centred text production, intercultural communication, conceptual transfer, ethics, risk, quality assurance and consultancy.

Above all, translator educators need to reach out to corporate communications and other related disciplines where the demand for high-quality human translation is most likely to continue and grow. As already mentioned, the MA at Zurich University of Applied Sciences has launched a series of authentic collaborative projects directly involving translation, interpreting and organisational communication students in the design and application of corporate communications concepts. Flanking courses on corporate communications, marketing and/or branding for translation students would prove equally beneficial, while familiarity with the essentials of intercultural communication and the cognitive semantics of conceptual transfer will reinforce student awareness of audience design and user-centred translation. We are currently exploring how, where and when these components might best be introduced into the curriculum.

The growing trend towards technologisation has caused an imbalance in translator education. By re-humanising professional translation and focussing on the adaptive, (co-)creative and consultative value-added potential of translators, we can empower them to develop the higher-autonomy, strategically oriented self-concept that both they and the organisations employing their services require.

FUTURE RESEARCH DIRECTIONS

The results of the survey of translation and communications professionals, backed by the insights from the pre-professional course evaluation data, have together produced preliminary results indicating considerable potential for an improved interaction between translators and corporate communications specialists. However, there are limitations to the objective reliability and validity of the findings. The students' course evaluations and feedback sessions could have been influenced by the phenomenon of social desirability, as the students were aware of the teachers' intention of encouraging the pre-professionals to get to know each other's field. Moreover, though the survey itself produced a relatively large set of data, the unequal numbers of responses between translators and project managers on the one hand, and corporate communications specialists on the other, pose issues of representativeness and robustness. Finally, it became apparent during the analysis of the data that some of the statements in the questionnaire to which the participants had to respond were potentially ambiguous or vague, particularly those about giving and receiving advice and feedback.

Additional research is therefore clearly required before any authoritative recommendations for translation practitioners and their education can be made. The most immediate step will be to analyse the complete survey data in greater detail, which should hopefully bring relations between the variables more clearly into focus, and to triangulate them with those of the semi-structured interviews conducted with freelance and staff translators. It is hoped that these will provide answers to at least some of the questions left open by the survey, first and foremost the discrepancy between awareness of communication strategies and access to communications plans. Further interviews with corporate communications specialists are also planned in order to identify important points of contact between the two professions that would allow for a more productive exchange of information.

As previously mentioned, we intend to follow up this pilot study with an international empirical project to survey, explore and enhance, from both a qualitative and an educational perspective, the processes and products of multilingual reputational text production by combining and refining relevant models from translation and corporate communications. The shortcomings of the present survey have already supplied useful insights into how the data-gathering instruments can be improved. It is anticipated that this, together with the prospect of addressing a wider public beyond Switzerland, will make for the collection of more robust data.

CONCLUSION

Despite the caveats mentioned above, the study reported in this contribution has furnished valuable knowledge about the current interplay between translation and corporate communications in Switzerland. With data from a large number of translation professionals, it has provided quite clear indicators of restricted access to communications plans and suggested confusion, vagueness and even ignorance

about the nature of corporate communication strategies. It has also revealed that, despite a healthy willingness and ability to ask questions, advisory feedforward processes between communications specialists and translators are largely inadequate and feedback flows from translators to clients, writers and project managers are highly restricted. The latter can be plausibly related to the survey results on translators' self-concept, which support the findings from extensive prior empirical research by showing that, despite a strong awareness of the duty to assist clients and adapt their texts, most translators in the survey do not necessarily see themselves in the mediatory, co-creative or advisory role that meaningful cooperation with corporate communications would imply. If further research serves to validate the results presented here, practitioners and theorists from both fields would do well to review and expand their models and processes for quality assurance and assessment, while educators and their institutions should set about promoting the extended self-concept on which translators' potentially new professional roles and perspectives are predicated.

REFERENCES

- Argenti, P. A. (2013). *Corporate communication* (6th ed.). New York: McGraw-Hill Education.
- Bayer-Hohenwarter, G. (2010). Comparing translational creativity scores of students and professionals: Flexible problem-solving and/or fluent routine behaviour? In S. Göpferich, F. Alves, & I. M. Mees (Eds.), *New approaches in translation process research* (pp. 83–111). Copenhagen: Samfundslitteratur Press.
- Bayer-Hohenwarter, G. (2011). Kreativität in populärwissenschaftlichen und instruktiven Texten im Vergleich: Kein Raum für Kreativität beim Übersetzen von instruktiven Texten? *trans-kom*, 4(1), 49-75.
- Canfora, C., & Ottmann, A. (2015). Risikomanagement für Übersetzungen. *trans-kom*, 8(2), 314-346. Retrieved April 10, 2017, from http://www.trans-kom.eu/bd08nr02/trans-kom_08_02_02_Canfora_Ottmann_Risiko.20151211.pdf
- Canon, M., & Marifian, E. (2013). Job polarization leaves middle-skilled workers out in the cold. *The Regional Economist*, 9-11. Retrieved June 9, 2017, from https://www.stlouisfed.org/~media/Files/PDFs/publications/pub_assets/pdf/re/2013/a/employment.pdf
- Carl, M., Bangalore, S., & Schaeffer, M. (Eds.). (2016). *New directions in empirical translation process research: Exploring the CRITTTPR-DB*. Cham: Springer International. doi:10.1007/978-3-319-20358-4
- Carl, M., Gutermuth, S., & Hansen-Schirra, S. (2015). Post-editing machine translation: Efficiency, strategies, and revision processes in professional translation settings. In A. Ferreira & J. W. Schwieter (Eds.), *Psycholinguistic and cognitive inquiries into translation and interpreting* (pp. 145–174). Amsterdam: Benjamins. doi:10.1075/btl.115.07car
- Chandler, D., & Munday, R. (2011). *A dictionary of media and communication*. Oxford, UK: Oxford University Press. doi:10.1093/acref/9780199568758.001.0001
- Cornelissen, J. (2014). *Corporate communication: A guide to theory and practice* (4th ed.). London: Sage.
- Dodds, R. A., & Smith, S. M. (1999). Fixation. In M. A. Runco & S. R. Pritzker (Eds.), *Encyclopedia of Creativity* (pp. 725–728). San Diego, CA: Academic Press.

Dvorkin, M. (2016). Jobs involving routine tasks aren't growing. *On the Economy, Federal Reserve Bank of St. Louis*. Retrieved June 9, 2017, from <https://www.stlouisfed.org/on-the-economy/2016/january/jobs-involving-routine-tasks-arent-growing>

Ehrensberger-Dow, M., & Massey, G. (2013). Indicators of translation competence: Translators' self-concepts and the translation of titles. *Journal of Writing Research*, 5(1), 103–131. doi:10.17239/jowr-2013.05.01.5

Ehrensberger-Dow, M., & Massey, G. (2014). Cognitive ergonomic issues in professional translation. In J. W. Schwieter & A. Ferreira (Eds.), *The development of translation competence: Theories and methodologies from psycholinguistics and cognitive science* (pp. 58–86). Newcastle: Cambridge Scholars.

Ehrensberger-Dow, M., & Massey, G. (2017). Socio-technical issues in professional translation practice. *Translation Spaces*, 6(1), 104–121. doi:10.1075/ts.6.1.06ehr

EMT Expert Group. (2009). *Competences for professional translators, experts in multilingual and multimedia communication*. Brussels: European Commission, Directorate-General for Translation. Retrieved October 30, 2016, from http://ec.europa.eu/dgs/translation/programmes/emt/key_documents/emt_competences_translators_en.pdf

EN 15038. (2006). *Translation services — service requirements*. Brussels: European Committee for Standardization.

European Commission. Directorate-General for Translation. (2012). *The status of the translation profession in the European Union. Final report*. Luxembourg: Publications Office of the European Union.

Feedback. (2017). In *Merriam-Webster.com*. Retrieved October 14, 2017, from <http://www.merriam-webster.com/dictionary/feedback>

Feedforward. (2017). In *Oxforddictionaries.com*. Retrieved October 14, 2017, from <https://en.oxforddictionaries.com/definition/feedforward>

Fiederer, R., & O'Brien, S. (2009). Quality and machine translation: A realistic objective? *Journal of Specialised Translation*, 11, 52–74.

Göpferich, S. (2008). *Translationsprozessforschung. Stand - Methoden - Perspektiven*. Tübingen: Narr.

Göpferich, S. (2012). Tracing strategic behaviour in translation processes: Translation novices, 4th-semester students and professional translators compared. In S. Hubscher-Davidson & M. Borodo (Eds.), *Global trends in translator and interpreter training* (pp. 240–266). London: Continuum.

Gouadec, D. (2007). *Translation as a profession*. Amsterdam: Benjamins. doi:10.1075/btl.73

Grace, K., Salvatier, J., Dafoe, A., Zhang, B., & Evans, O. (2017). *When will AI exceed human performance? Evidence from AI experts*. Retrieved June 9, 2017, from <https://arxiv.org/pdf/1705.08807.pdf>

Guerberof Arenas, A. (2009). Productivity and quality in the post-editing of outputs from translation memories and machine translation. *Localisation Focus*, 7(1), 11–21.

Guerberof Arenas, A. (2013). What do professional translators think about post-editing? *Journal of Specialised Translation*, 19, 75–95.

- Guerberof Arenas, A. (2014a). The role of professional experience in post-editing from a quality and productivity perspective. In S. O'Brien, L. Winther Balling, M. Carl, M. Simard, & L. Specia (Eds.), *Post-editing of machine translation: Processes and applications* (pp. 51–76). Newcastle: Cambridge Scholars.
- Guerberof Arenas, A. (2014b). Correlations between productivity and quality when post-editing in a professional context. *Machine Translation*, 28(3-4), 165–186. doi:10.1007/10590-014-9155-y
- Hagemann, S. (2016). (Non-)professional, authentic projects? Why terminology matters. In D. Kiraly, L. R  th, C. Canfora & A. Cnyrim (Eds.), *Towards authentic experiential learning in translator education* (pp. 33-51). G  ttingen: V&R unipress/Mainz University Press.
- Hofmann, S. (2012). *Prozessgest  tztes   bersetzen. Vom funktionsorientierten   bersetzungsprozess zum GP-Modell f  r die Dienstleistung   bersetzen*. Lichtenberg: Harland Media.
- Holtzhausen, D., & Zerfass, A. (2015). *The Routledge handbook of strategic communication*. New York: Routledge.
- H  nig, H. G., & Kussmaul, P. (1982). *Strategie der   bersetzung. Ein Lehr- und Arbeitsbuch*. T  bingen: Narr.
- House, J. (2015). *Translation quality assessment: Past and present*. London: Routledge.
- Huck-Sandhu, S. (2015). Internationale PR-Arbeit. In G. Bentele, R. Fr  hlich, & P. Szyszka (Eds.), *Handbuch Public Relations. Wissenschaftliche Grundlagen und berufliches Handeln* (pp. 757–772). Wiesbaden: Springer VS.
- Huck-Sandhu, S. (2016). Internationale Unternehmenskommunikation. In C. Mast (Ed.), *Unternehmenskommunikation. Ein Leitfaden* (pp. 423–440). Konstanz: UVK.
- ISO 17100. (2015). *Translation services — Requirements for translation services*. Geneva: ISO.
- Katan, D. (2011). Occupation or profession: A survey of the translators' world. In R. Sela-Sheffy & M. Shlesinger (Eds.), *Profession, identity and status: Translators and interpreters as an occupational group* (pp. 65–88). Amsterdam: Benjamins. doi:10.1075/bct.32.06kat
- Katan, D. (2013). Cultural mediation. In Y. Gambier & D. Van Doorslaer (Eds.), *Handbook of translation studies*, 4 (pp. 84–91). Amsterdam: Benjamins. doi:10.1075/hts.4.int5
- Katan, D. (2016). Translation at the cross-roads: Time for the transcreational turn? *Perspectives*, 24(3), 365–381. doi:10.1080/0907676X.2015.1016049
- Kiraly, D. (1997). Think-aloud protocols and the construction of a professional translator's self-concept. In J. H. Danks, G. M. Shreve, S. B. Fountain, & M. K. McBeath (Eds.), *Cognitive processes in translation and interpreting* (pp. 137–160). Thousand Oaks, CA: Sage.
- Kiraly, D., R  th, L., Canfora, C., & Cnyrim, A. (Eds.). (2016). *Towards authentic experiential learning in translator education*. G  ttingen: V&R unipress/Mainz University Press.
- Koponen, M. (2016). Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *Journal of Specialised Translation*, 25, 131–148.

Kranich, S., Becher, V., Höder, S., & House, J. (Eds.). (2011). *Multilingual discourse production: Diachronic and synchronic perspectives*. Amsterdam: Benjamins. doi:10.1075/hsm.12

Kurz, C. (2009). Translatorisches Qualitätsmanagement als verantwortungsvolles Handeln. *Lebende Sprachen*, 54(4), 146–155. doi:10.1515/les.2009.038

Läubli, S., Fishel, M., Massey, G., Ehrensberger-Dow, M., & Volk, M. (2013). Assessing post-editing efficiency in a realistic translation environment. In *Proceedings of the XIV MT summit workshop on post-editing technology and practice* (pp. 83–91). Allschwil: European Association for Machine Translation.

Liddicoat, A. J. (2016). Intercultural mediation, intercultural communication and translation. *Perspectives*, 24(3), 354–364. doi:10.1080/0907676X.2014.980279

Massardo, I., van der Meer, J., & Khalilov, M. (2016). *Translation technology landscape report. September 2016*. De Rijp: TAUS.

Mast, C. (2016). *Unternehmenskommunikation. Ein Leitfadens* (6th ed.). Konstanz: UVK.

Mellinger, C. D., & Shreve, G. M. (2016). Match evaluation and over-editing in a translation memory environment. In R. Muñoz Martín (Ed.), *Reembedding translation process research* (pp. 131–148). Amsterdam: Benjamins. doi:10.1075/btl.128.07mel

O'Brien, S. (2012). Translation as human-computer interaction. *Translation Spaces*, 1, 101–122. doi:10.1075/ts.1.05obr

O'Brien, S., Winther Balling, L., Carl, M., Simard, M., & Specia, L. (Eds.). (2014). *Post-editing of machine translation: Processes and applications*. Newcastle: Cambridge Scholars.

Olohan, M. (2011). Translators and translation technology: The dance of agency. *Translation Studies*, 4(3), 342–357. doi:10.1080/14781700.2011.589656

Perrin, D., & Stücheli-Herlach, P. (2013). Schreiben mit System: Texte planen, entwerfen und verbessern für die medienkonvergente PR. In D. Perrin & P. Stücheli-Herlach (Eds.), *Schreiben mit System* (pp. 15–38). Wiesbaden: Springer VS.

Prassl, F. (2010). Translators' decision-making processes in research and knowledge integration. In S. Göpferich, F. Alves, & I. M. Mees (Eds.), *New approaches in translation process research* (pp. 57–81). Copenhagen: Samfundslitteratur Press.

Pym, A., Orrego-Carmona, D., & Torres-Simón, E. (2015). Status and technology in the professionalization of translators. Market disorder and the return of hierarchy. *Journal of Specialised Translation*, 25, 33–53.

Quirke, B. (2008). *Making the connections: Using internal communication to turn strategy into action*. Wallingford: Gower.

Rasinger, S. M. (2013). *Quantitative research in linguistics: An introduction* (2nd ed.). London: Bloomsbury.

Risku, H. (2010). A cognitive scientific view on technical communication and translation: Do embodiment and situatedness really make a difference? *Target*, 22(1), 94–111. doi:10.1075/target.22.1.06ris

Saldanha, G., & O'Brien, S. (2014). *Research methodologies in translation studies*. London: Routledge.

Schmid, B., & Lyczek, B. (2006). Die Rolle der Kommunikation in der Wertschöpfung der Unternehmung. In B. Schmid & B. Lyczek (Eds.), *Unternehmenskommunikation. Kommunikationsmanagement aus Sicht der Unternehmensführung* (pp. 3–146). Wiesbaden: Gabler. doi:10.1007/978-3-8349-9126-3_1

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., & Norouzi, M. (2016). *Google's neural machine translation system: Bridging the gap between human and machine translation*. Retrieved June 9, 2017, from <https://arxiv.org/pdf/1609.08144.pdf>

ENDNOTES

- ¹ The title of the specialisation may appear misleading in the English-speaking world, as its objectives and content are primarily intended to educate students in corporate communications.
- ² A licensed version of the LimeSurvey online tool was used; see <https://www.limesurvey.org/>
- ³ According to the ranking conducted by the Swiss weekly *Handelszeitung* in 2016.
- ⁴ The last response was provided by a participant who was in charge of a team of translators.
- ⁵ e.g. the Lionbridge company CLS Communication; see <https://www.cls-communication.com/en>

APPENDIX 1

Survey for Translators and Translation Project Managers

1. First and last name (optional):
2. Country of residence (optional):
3. E-mail address (optional):
4. Years of professional experience to the nearest year:
5. Please select the statement below that most closely describes your work profile:

I am a translator who mainly works freelance.

I am a translator who mainly works as a staff translator for government/public institution or NGO.

I am a translator who mainly works as a staff translator for a commercial organisation.

I am a translation project manager who works for a government/public institution or NGO.

I am a translation project manager who works for a commercial organisation.

6. As a translator, I see my principal role(s) as

remaining faithful to the source-text writers' intentions. / mediating between language and cultural groups. / providing feedback on source texts and/or assignments. / transferring the main messages of the source text. / co-creating texts with clients and/or source-text authors. / meeting document specifications and/or demands. / fulfilling my clients' wishes and/or requirements. / meeting project management standards and/or deadlines. / transforming source texts for my target audiences. / helping clients communicate messages and/or intentions. / other ...

7. I am aware of the communication strategy of the client/organisation(s) I work for.

Please choose only one of the following: I agree fully / I agree mostly / I agree partly / I do not agree / Not applicable

8. I have access to the communications plan (i.e. the written form of the communication strategy) of the client/organisation(s) I work for.

Please choose only one of the following: I agree fully / I agree mostly / I agree partly / I do not agree / Not applicable

9. I receive general advice on how to contribute to the communications objectives of the client/organisation(s) I work for.

Please choose only one of the following: I agree fully / I agree mostly / I agree partly / I do not agree / Not applicable

Please specify in brief the sort of advice you have received.

10. I receive specific linguistic advice on how to contribute to the communications objectives of the client/organisation(s) I work for.

Please choose only one of the following: I agree fully / I agree mostly / I agree partly / I do not agree / Not applicable

Please specify in brief the sort of advice you have received.

11. I can ask the client/organisation(s) directly any questions I have about the communication strategy.

Please choose only one of the following: I agree fully / I agree mostly / I agree partly / I do not agree / Not applicable

12. (*Only for freelance and staff translators*) I can ask translation project managers any questions I have about the communication strategy.

Please choose only one of the following: I agree fully / I agree mostly / I agree partly / I do not agree / Not applicable

13. I receive feedback on the strategic adequacy of our/my texts from the client/organisation(s).

Please choose only one of the following: I agree fully / I agree mostly / I agree partly / I do not agree / Not applicable

14. I receive feedback on the strategic adequacy of our/my texts from source-text writers.

Please choose only one of the following: I agree fully / I agree mostly / I agree partly / I do not agree / Not applicable

15. (*Only for freelance and staff translators*) I receive feedback on the strategic adequacy of our/my texts from translation project managers.

Please choose only one of the following: I agree fully / I agree mostly / I agree partly / I do not agree / Not applicable

16. (*Only for translation project managers*) I receive feedback on the strategic adequacy of our/my texts from translators.

Please choose only one of the following: I agree fully / I agree mostly / I agree partly / I do not agree / Not applicable

17. (*Only for staff translators and translation project managers*) I work for an organisation that operates mainly

Choose one of the following answers: in Switzerland / internationally

18. (*Only for staff translators and translation project managers*) I work for an organisation in the following industry/sector:

Choose one of the following answers: Manufacturing industry / Trade / Financial services, insurance / Energy / Building, housing / Tourism, hospitality / Healthcare / Education, research / Transportation / Media / Information, communication / Public administration / Other ...

APPENDIX 2

Survey for Communications Specialists

1. First and last name (optional):
2. Country of residence (optional):
3. E-mail address (optional):
4. Years of professional experience to the nearest year:
5. Please select the statement below that most closely describes your work profile:

I work as a corporate communications professional who frequently requests translations.

I work as a corporate communications professional who writes texts for external and/or internal PR.

6. I am aware of the communication strategy of the organisation I work for.

Please choose only one of the following: I agree fully / I agree mostly / I agree partly / I do not agree / Not applicable

7. I have access to the communications plan (i.e. the written form of the communication strategy) of the organisation I work for.

Please choose only one of the following: I agree fully / I agree mostly / I agree partly / I do not agree / Not applicable

8. (*Only for corporate communications professionals who write PR texts*) I receive general advice on how to contribute to the communications objectives of the organisation I work for.

Please choose only one of the following: I agree fully / I agree mostly / I agree partly / I do not agree / Not applicable

Please specify in brief the sort of advice you have received.

9. I give general advice to translation service providers on how to contribute to the communications objectives of the organisation I work for.

Please choose only one of the following: I agree fully / I agree mostly / I agree partly / I do not agree / Not applicable

Please specify in brief the sort of advice you give.

10. (*Only for corporate communications professionals who write PR texts*) I receive specific linguistic advice on how to contribute to the communications objectives of the organisation I work for.

Please choose only one of the following: I agree fully / I agree mostly / I agree partly / I do not agree / Not applicable

Please specify in brief the sort of advice you have received.

11. I give specific linguistic advice to translation service providers on how to contribute to the communications objectives of the organisation I work for.

Please choose only one of the following: I agree fully / I agree mostly / I agree partly / I do not agree / Not applicable

Please specify in brief the sort of advice you give.

12. (*Only for corporate communications professionals who write PR texts*) I can ask the organisation directly any questions I have about the communication strategy.

Please choose only one of the following: I agree fully / I agree mostly / I agree partly / I do not agree / Not applicable

Please specify in brief the sort of advice you have received.

13. (*Only for corporate communications professionals who write PR texts*) I receive feedback on the strategic adequacy of our/my texts from the organisation.

Please choose only one of the following: I agree fully / I agree mostly / I agree partly / I do not agree / Not applicable.

Please specify in brief the sort of advice you give.

14. I receive feedback on the strategic adequacy of our/my texts from translation project managers.

Please choose only one of the following: I agree fully / I agree mostly / I agree partly / I do not agree / Not applicable

15. I receive feedback on the strategic adequacy of our/my texts from translators.

Quality Assurance in Translation and Corporate Communications

Please choose only one of the following: I agree fully / I agree mostly / I agree partly / I do not agree / Not applicable

16. I work for an organisation that operates mainly

Choose one of the following answers: in Switzerland. / internationally.

17. I work for an organisation in the following industry/sector:

Choose one of the following answers: Manufacturing industry / Trade / Financial services, insurance / Energy / Building, housing / Tourism, hospitality / Healthcare / Education, research / Transportation / Media / Information, communication / Public administration / Other ...

Section 2

Translation Quality Assessment in Interpreting and Audio- Visual Translation

Chapter 4

Conceptualizing and Operationalizing a Formative Assessment Model for English–Chinese Consecutive Interpreting: A Case Study in an Undergraduate Interpreting Course

Chao Han
Southwest University, China

ABSTRACT

Formative assessment has been increasingly used by interpreter trainers and educators to promote student learning. Different forms of formative assessment have been practiced and reported in interpreting literature. However, a critical review of current practices reported in literature suggests that a longitudinally designed formative assessment model that harnesses the synergistic potential of self, peer, and teacher assessment seems to be lacking. This chapter therefore aims to provide a detailed account of how an inclusive formative assessment model was conceptualized and operationalized for an undergraduate-level English–Chinese consecutive interpreting course and how students and the teacher perceived the assessment model. Based on the students' evaluation and the teacher's reflection, the chapter highlights good practices that contribute to effective formative assessment, discusses potential problems, proposes possible solutions, and suggests future trends in implementing and researching formative assessment in interpreter training and education.

DOI: 10.4018/978-1-5225-5225-3.ch004

INTRODUCTION

Testing and assessment play an important role in interpreter education. Previous literature has documented how potential candidates for a given interpreting program could be tested and selected in an admission test (e.g., Russo, 2014); how formative assessment could be conducted to diagnose strengths and weaknesses of student interpreters (e.g., Lee, 2016), as well as how trainees could be tested so as to exit a program (e.g., Sawyer 2004) and be certified as junior professional interpreters (e.g., Han & Slatyer, 2016). To some extent, admission testing functions as a gate-keeping mechanism for interpreting programs, while exit testing aims to ascertain the achievement of trainee interpreters and evaluate their job readiness for the interpreting market. Both testing practices serve a regulatory purpose, controlling accessibility to either educational resources or the interpreting market. Contrary to admission and exit testing, formative assessment is intended to provide feedback on students' performance to improve and accelerate learning (Sadler, 1998). Over the years, the value of formative assessment has been increasingly recognized by interpreter trainers and educators, and different forms of formative assessment have thus been trialled in interpreter training programs. For instance, self-assessment is generally believed to help student interpreters develop reflective thinking and foster learner autonomy (Witter-Merithew, Taylor, & Johnson, 2001). As such, many interpreting researchers regard self-assessment as an integral part in interpreter education (Iaroslavschi, 2011; Sandrelli, 2015). Peer assessment is also valued as an important form of formative assessment and is practiced in interpreting courses (Lee, 2016; Lim, 2013). Despite the growing recognition and practice of formative assessment in the interpreting classroom, it seems that few researchers have attempted to design a formative assessment model that is inclusive of self, peer and teacher assessment (see otherwise Fowler, 2007), that is implemented on a regular basis throughout an academic semester, and that is subject to both students' and teacher's evaluation. Against this background, the present chapter aims to provide a detailed account of how a formative assessment model was conceptualized and operationalized for an undergraduate-level English-Chinese consecutive interpreting course, and how students and the teacher perceived the assessment model.

FORMATIVE ASSESSMENT IN HIGHER EDUCATION

This section provides a brief theoretical review of formative assessment. It is based on the literature of higher education. Particularly, definitions of formative assessment are provided, its key characteristics highlighted, and implementation strategies discussed. In the 1960s, Scriven (1967) first used the term "formative evaluation", but its usage was primarily related to the improvement of curriculum. It was Bloom, Hastings, and Madaus (1971) that first used "formative evaluation" in connection with promoting student learning through assessment. Bell and Cowie (2001, p. 537) define formative assessment as "the process used by teachers and students to recognise and respond to student learning in order to enhance that learning, during the learning." Bell and Cowie's definition brings to the fore five characteristics of formative assessment that set it apart from other types of assessment (e.g., summative assessment). The first characteristic is that formative assessment is an on-going and progressive process, not a one-off and static activity. In this sense, (planned or unplanned) formative assessment needs to be carried out longitudinally over an extended period of time. Secondly, formative assessment happens during the process of learning, not *before* or *after* learning. This characteristic distinguishes formative assessment from placement assessment (conducted before actual learning) and summative assessment (conducted at

Conceptualizing and Operationalizing a Formative Assessment Model

the end of an instructional period). Thirdly, the agents or stakeholders of formative assessment include both teachers and students. Traditionally, teachers are believed to be responsible for creating an effective learning environment and for promoting student learning within that environment. However, since students also have the responsibility for their own learning, it is also incumbent on students to play an important role in promoting their learning. Fourthly, both stakeholders need to recognize and respond to learning. That is, teachers and students should elicit evidence of learning, and take relevant actions in response to lack of progress. Lastly, the purpose of formative assessment is to enhance student learning. In this sense, formative assessment is essentially assessment *for* learning, not *of* learning.

Among the characteristics of formative assessment, the most important one is, arguably, the provision of feedback. Sadler (1998, p. 79) observes that “incorporating feedback is surely as fundamental a characteristic of responsible and responsive learning systems as having a teacher at all.” Specifically, feedback is defined by Ramaprasad (1983, p. 4) as “information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way.” Based on Sadler’s definition, a feedback system consists of four elements: 1) data on the actual level of a given attribute; 2) data on the reference or desired level of the attribute; 3) a mechanism for comparing the two levels and generating information about the gap between the two levels, and 4) a mechanism by which the information can be used to alter the gap (Black & William, 1998, p. 48). In a sense, formative assessment aims to identify the gap between a learner’s actual level and the desired level as well as find out ways to narrow that gap.

To maximize the potential benefits of formative assessment, William and Thompson (2007) propose five key strategies: 1) clarifying and sharing learning intentions and criteria for success; 2) conducting classroom discussions and other learning tasks that elicit evidence of student learning and understanding; 3) providing feedback that motivates students to improve; 4) activating students as instructional resources for their peers, and 5) activating students as the owners of their own learning. The first strategy implies that teachers need to make assessment rationales and criteria clear and transparent to students. The second strategy requires the provision of opportunities to students so that they can demonstrate outcomes of learning. Regarding the third strategy, it cannot simply be assumed that when feedback is provided to students, they will effectively use and engage with it. Students may need to be trained in how to interpret feedback (be it from teachers or peers) and how to relate the feedback to their performance. The fourth and fifth strategies imply that both self and peer assessment can be fruitfully combined to enrich the formative assessment experience.

FORMATIVE ASSESSMENT IN INTERPRETER TRAINING AND EDUCATION

The use of formative assessment in the interpreting classroom has gained momentum over the years (Arumí Ribas, 2010; Sawyer, 2004). This is largely because interpreting educators have recognized the benefits of formative assessment, such as generating feedback to student learning, revealing strengths and weaknesses of trainee interpreters, as well as fostering learner autonomy and metacognitive awareness (Sandrelli, 2015; Witter-Merithew et al. 2001).

According to the extant interpreting literature, formative assessment has been operationalized in four forms: 1) self-assessment; 2) peer assessment; 3) a combination of self and peer assessment, and 4) student portfolio. Among these forms, the practice of self-assessment has been mostly reported in the literature (e.g., Han & Riazi, 2018; Iaroslavschi, 2011; Ibrahim-González & Noordin, 2012; Lee,

2011; Moser-Mercer, 1996; Russo, 1995), in which interpreter trainers and educators discuss how self-assessment can be conducted in the interpreting classroom. Notably, several authors provide detailed accounts of how self-assessment models are designed and implemented in various training contexts (see Choi, 2006; Lee, 2005; Postigo Pinazo, 2008; Schjoldager, 1996; Witter-Merithew et al. 2001). For instance, Witter-Merithew et al. (2001) illustrate a guided self-assessment model applied to sign language interpreters enrolled in the Educational Interpreting Certificate Program (EICP) in the United States. The self-assessment model is characterized as an on-going process that occurs at planned intervals throughout the EICP and consists of five main steps: 1) analysis of source text; 2) videotape production of signed/interpreted samples; 3) transcription of signed/interpreted performance; 4) analysis of performance, and 5) formulation of ideas for self-improvement. Choi (2006) also describes a metacognitive model concerning self-assessment of Korean-English consecutive interpretation for novice interpreting students. The model features five stages: 1) student self-evaluation; 2) problem-finding; 3) prioritization of critical issues identified; 4) remedial practice, and 5) reevaluation and monitoring.

Several common characteristics emerge from these self-assessment models. Firstly, while some models focus entirely on assessing the quality of interpreting product (Choi, 2006; Lee, 2005; Schjoldager, 1996), others cast light on both the interpreting process (i.e., note-taking) and the product (i.e., renditions) (Postigo Pinazo, 2008). Secondly, when self-assessing their performances, students are asked to identify strengths and weaknesses of their interpretation. In some cases, they also need to formulate specific plans and strategies to address identified problems (Lee, 2005; Witter-Merithew et al., 2001). Thirdly, although the self-assessment models foreground students as self-assessors and autonomous learners, they invariably emphasize the teachers' role as organizers, facilitators and advisors, who can provide valuable feedback to complement students' self-reflection. Notably, in the model by Witter-Merithew et al. (2001), three types of feedback can be provided by instructors, including: 1) written narrative feedback; 2) videotaped feedback, and 3) face-to-face feedback, to meet students' specific needs. In general, this body of literature produces some positive evidence supporting the effectiveness of self-assessment. It is reported that self-assessment helps students gain an objective understanding of their performances (Lee, 2005), nurture a positive attitude on learning and improve their academic performance (Postigo Pinazo, 2008).

When it comes to using formative peer assessment in interpreter training, relevant literature is few and far between. One exception is Lee (2016), who describes a peer assessment model implemented in an English-Korean consecutive interpreting course for two successive semesters. One of the noteworthy characteristics of the model is the provision of training sessions by the teacher, in which the students were trained in how to use a rating scale in peer assessment. Also of note is that peer assessment is conducted collectively during regular class time, in which the entire set of interpretation recordings is played twice, but in a different sequence. One additional characteristic is that the teacher elicits students' opinions on peer assessment by asking them to write reflective journals or conduct personal interviews. Based on the students' feedback, Lee (2016) observes that although the students reported difficulty in applying criteria and suffered from rating fatigue, they felt empowered, demonstrated responsibility, and regarded peer assessment as being beneficial and reliable. Another article related to peer assessment in interpreter training is Lim's (2013), which, however, does not describe an assessment model per se, but relies on a questionnaire to find out students' experiences and perceptions of peer assessment in a major interpreter training institution in South Korea. All 96 questionnaire respondents commented that they had made peer assessments of other students in class; 60% of the students said that they had received some training from instructors regarding how to conduct peer assessment. These findings may suggest

Conceptualizing and Operationalizing a Formative Assessment Model

that although (in)formal peer assessment is widely practiced in interpreter education, it is under-reported in the literature.

Self and peer assessment are also used in tandem by a few interpreter trainers. In applying self and peer assessment in interpreted role-plays, Fowler (2007) highlights three aspects of her assessment model: 1) providing relevant training to students; 2) generating multi-perspective formative feedback, and 3) ritualizing feedback exercises. Similar to Lee (2016), Fowler (2007) emphasizes the provision of proper training to students before the actual assessment. The training aims to inform students of the rationale of self and peer assessment, explicate assessment criteria to be used, and explain how to receive and give oral and written feedback to peers. In addition, formative feedback from multiple sources (i.e., students themselves, their peers and possibly their tutors) is generated, which helps heighten students' awareness of their behavior as interpreters and facilitate a genuine and informed dialogue about students' performance. Lastly, feedback exercise (i.e., students' writing up observations and awarding grades to their peers) is set as a regular assignment. This measure boosts the status of the exercise and requires students' commitment. Another piece of writing pertinent to self and peer assessment is Hartley, Mason, Peng, and Perez (2003). However, the article primarily concerns the design of comprehensive and applicable guidelines that can be used for self-monitoring and peer assessment. As such, there is no detailed description of how self and peer assessment is actually implemented.

Finally, a number of interpreter educators use student portfolio as a formative assessment tool (Arumí Ribas, 2010; Sawyer, 2004). In general, student portfolio is a compilation of important documents (print and/or digital), which records each student's performance and achievement during the learning process. The portfolio could include a wide range of materials evidencing both strengths and weaknesses of a student's academic performance. For example, useful materials could include statement of personal goals, video and audiotapes of student's work, samples of written work, self-assessment statements, and peer review records. Arguably, one of the most significant aspects of the portfolio is that students need to reflect on the samples included in the portfolio. Through constant self-reflection, students develop a sense of self-awareness and improve self-regulation. Applying a model of student portfolio in a German-Spanish/Catalan simultaneous interpreting classroom, Arumí Ribas (2010) provides a guide for drawing up the portfolio to students, which includes course syllabus, a personal goal sheet, weekly interpreting performance self-corrections, a CD with digital oral interpretation files, in-class correction diary, sheets featuring results from tutorials with instructors' oral presentations given in class, a final reflective report, and exam results. According to a subsequent questionnaire completed by the students, the portfolio was regarded as a good alternative to traditional exams and a good method of continuous assessment.

As can be seen from the literature review above, formative assessment operationalized in interpreter education features different practices, and is perceived by students as being beneficial and useful overall. Despite the positive results, there is still some room for further improvement. Firstly, a series of formative assessments could be planned and conducted by teachers throughout a curriculum so as to generate feedback and scaffold students' learning at different junctures of study. It is thus a pity that many of the previous authors conducted planned formative assessment only once in a given semester (Lee, 2005; Lee, 2016). Secondly, students could be involved somehow in devising assessment criteria so as to improve assessment transparency, sharpen their awareness, and increase their responsibility for learning (Orsmond, Merry, & Reiling, 2000; William & Thompson, 2007). Boud (1995) argues that if students only evaluate their or their peers' performance without concomitant involvement in establishing criteria, such practice should be referred to as self or peer grading, not as self or peer assessment. Thirdly, although formative assessment could take different forms, an assessment model that incorporates self, peer and teacher as-

assessment is arguably more likely to help students gain an in-depth and multi-perspective understanding of their own performance. Previous literature mostly draws on a single form of formative assessment (see Lee, 2016; Postigo Pinazo, 2008), without harnessing their synergistic potential.

Based on the above considerations, the author describes in the remainder of this chapter the conceptualization and operationalization of a formative assessment model for an undergraduate-level consecutive interpreting course. The author also reports students' evaluation of the model, foregrounds implementation problems from the teacher's perspective, and proposes possible solutions.

BACKGROUND TO THE STUDY

In the autumn semester of the 2016-2017 academic year (i.e., semester 7), the author convened and taught an advanced-level course of *Interpreting Practicum* in a national university located in Southwest China. The 18-week course aimed to provide ample opportunities for senior students in the school of foreign languages (occasionally, for capable and inspiring students from other colleges) to practice and enhance their English-Chinese consecutive interpreting skills. Each week, students met with the author once for about two hours. Normally, before registering in this course, students need to participate in three preceding courses, i.e., *Introductory*, *Intermediary* and *Advanced-level Interpreting*. In these courses, students have learnt a wide range of basic skills and techniques involved in consecutive interpreting (e.g., memorization, text analysis, analytic listening, note-taking). In *Interpreting Practicum*, students refine relevant skills by practicing interpreting extensively on a diverse range of non-technical topics. It is envisaged that upon graduation students would be able to perform consecutive interpreting for daily communicative events that do not require technical and specialized knowledge. It is also expected that the course would lay a foundation for those students who want to pursue postgraduate-level interpreter training.

For the autumn semester, a total of 41 fourth-year undergraduate students registered the course and were placed in two parallel classes¹. Enrolled in a Bachelor-of-Arts program, all students specialized in English-Chinese translation. They had studied consecutive interpreting for three consecutive semesters. Averaged at 21 years, 36 students were female (88%) and the rest of them were male (12%). They spoke Mandarin Chinese as mother tongue and had learnt English as a foreign language. The working languages were thus Mandarin Chinese (A language) and English (B language). It is worth noting that this group of students had practiced self-assessment in consecutive interpretation in the previous semester (i.e., semester 6), but had no experience of peer assessment or any combination of self, peer and teacher assessment.

A key driver of the formative assessment in the interpreting course is the university's teaching policy that focuses on a learner-centered, process-oriented approach to student learning. Teachers are encouraged to enable students to assume more responsibility for their learning, to scaffold students' learning at different stages of study, and to provide evidence of student progress to complement results of summative assessment. A longitudinally-designed formative assessment model is therefore an ideal candidate to engage with students and generate such evidence.

CONCEPTUALIZATION OF A FORMATIVE ASSESSMENT MODEL FOR CONSECUTIVE INTERPRETING

Based on the literature review and against the specific background, the author/teacher conceptualized four general characteristics of the formative assessment model. In general, these characteristics were largely in line with those highlighted by Bell and Cowie (2001). The only potential difference is that the current assessment model laid emphasis on assessment transparency in terms of whether assessment criteria and rating procedures are shared among all stakeholders. Firstly, formative assessment needed to be an on-going exercise carefully planned to span over the whole semester (Witter-Merithew et al., 2001). In other words, the assessment model should be based on a longitudinal design. Secondly, formative assessment should be a transparent practice. All stakeholders involved in the assessment, primarily students and teachers, should be well informed of assessment rationales, procedures and methods (Lee, 2016; Orsmond et al., 2000; William & Thompson, 2007). Particularly, they needed to know what assessment criteria were going to be used and how an interpretation was going to be evaluated. To establish transparency, some preparatory work needed to be done, even before formative assessment was officially conducted. Thirdly, the model needed to be inclusive, capitalizing on various forms of formative assessment (Fowler, 2007; Hartley et al., 2003). Specifically, an assessment model that leverages the respective strength of self, peer and teacher assessment was developed for a trial run. Fourthly, the model should place emphasis on the value of formative feedback to enhance student learning (Fowler, 2007; Sadler, 1998; William & Thompson, 2007; Witter-Merithew et al., 2001). All parties involved in the assessment needed to provide both quantitative and qualitative feedback so that a multi-perspective understanding of students' performance could be gained. These general conceptualizations guided the design and operationalization of the formative assessment model.

OPERATIONALIZATION OF THE FORMATIVE ASSESSMENT MODEL

This section provides a detailed account of how the author operationalized the formative assessment model, focusing on nine major aspects of the model: 1) rationalizing formative assessment, 2) choosing time points, 3) designing and developing interpreting tasks, 4) administering interpretation assessments, 5) designing and using performance assessment sheets, 6) providing assessment training, 7) implementing self and peer assessment, 8) carrying out teacher assessment, and 9) surveying students' satisfaction level.

Rationalizing Formative Assessment

Before the semester officially started, the author wrote a group email to all students registering the course, suggesting the possibility of conducting formative assessment (primarily self and peer assessment) in the semester. In the email, the author also explained to the students how self and peer assessment could facilitate their learning and shared with them some positive experience reported in the relevant literature. In the first class in week 1, the author discussed how the course would be managed and provided further rationales for implementing formative assessment.

Time Points of Formative Assessment

In the 18-week semester, the author planned and organized three formative assessments, respectively in week 5, week 10 and week 15, before the summative assessment was conducted in week 18. In other words, formative assessment was held approximately every four weeks. The four-week interval ensured that formative assessment could span the semester evenly and that the students could have sufficient time between assessments to reflect, practice and improve. The frequency of assessment was also deemed reasonable. If only one or two assessments were conducted, the opportunities for the students to develop the habit of self-reflection and self-regulation would be limited. But if the assessments were organized too frequently (e.g., five or six times a semester), the students would have to spend a disproportionate amount of time on self and peer assessment, without engaging in substantive practice in their spare time.

Tasks Used in Formative Assessment

Each formative assessment consisted of one English speech and one Chinese speech. The students were required to interpret both speeches. As a result, the author needed to prepare a total of six speeches/tasks for the assessments, three for each source language. To standardize the task development, the author followed a set of general guidelines tailor-made for the assessment (for similar guidelines, see Han & Riazi 2017). Firstly, both English and Chinese speeches to be used in the assessments must be based on authentic speeches that pertain to non-technical and non-specialized topics. Secondly, source-language texts must be recorded by native speakers. Thirdly, source-language texts should be revised and adapted to achieve approximately equal length. Specifically, in contrast to other practices in which words are used as a basic measurement unit, the author chose syllables as a basic unit to quantify text length (also see Mead, 2005; Pöchhacker, 1993). This was largely because the number of syllables as an indicator of text length compares well across languages, especially for interpreting in which two drastically different languages may be involved (e.g., English and Mandarin Chinese). In this case, one Chinese word/character is a syllable, while an English word usually has two or more syllables. Resultantly, if measured in words, a Chinese text of a given length usually has much fewer syllables than an English text of the same length. Therefore, using syllables to measure text length turns out to be more precise in cross-language comparison. Fourthly, a source-language speech should be segmented into multiple shorter sections so that the students can interpret section by section. Fifthly, the total duration of source-language speeches should range from two and a half minutes to three minutes. Sixthly, the delivery speed of source-language speeches should be about 160 – 170 syllables per minute or 100 – 110 words per minute. Past experience suggests that such rates of speech delivery generally represent a suitable range of speech rate for undergraduate students to practice note-taking and interpreting. Lastly, a background information sheet for each speech (including a short note on the topic and several key words) should be prepared so that it can be provided to students before the assessments to help in-advance preparation.

Interpretation Assessment

As has been said, three formative assessments were conducted in weeks 5, 10 and 15. On each occasion, the author sent the background information sheet to the students to help them familiarize with relevant topics one or two days before the actual assessment. The students performed interpreting for one English speech and one Chinese speech in an open-plan language lab equipped with acoustic systems tailored for

interpreter training. Every student had an electronic interface on which they could watch videos played by the teacher and a sound console through which they could adjust input sound and record their voices. For each formative assessment, the 41 students were randomly split into four groups, each consisting of about 10 students, so that each group went to the language lab by turns. Although the lab could house 41 students, it was considered that too many students performing interpreting together would be a distraction for some students. All students' performances were audio-recorded and saved as MP3 files for later evaluation.

Performance Assessment Sheet

The author designed and used a descriptor-referenced, scale-based performance assessment sheet in the formative assessments (for a copy of the sheet, see Table 1). Four aspects of the performance assessment sheet need to be highlighted and explained: 1) assessment criteria; 2) format of the rating scale; 3) scalar descriptors, and 4) formative feedback.

In identifying assessment criteria, the author engaged the students in the process. In week 1 the author asked the students to think about what criteria could be used to evaluate their interpreting performance in the course. Relevant literature was also provided to the students (e.g., Han, 2015, 2016; Lee, 2008; Wang, Napier, Goswell, & Carmichael, 2015), so that they could consult past studies and be acquainted with previous discussions. In week 2, the author invited the students to speak up their minds in an interactive discussion session. Before the discussion, the author made it clear that once assessment criteria were identified and established, they were to be used in all types of assessment throughout the semester. After a lively discussion, three quality dimensions of interpretation emerged as assessment criteria: 1) content, 2) delivery and 3) language use. Specifically, the content aspect of interpreting performance concerns to what extent source-language information is successfully translated; the delivery aspect concerns to what extent disfluency features, such as filled pauses, hesitations and repairs, are present in the interpretation; and language use concerns to what extent target-language expressions are natural to a native English or Chinese speaker. These assessment criteria were then termed as: *information completeness*, *fluency of delivery*, and *target language quality*. In addition, a majority of the students voted for a weighting scheme which gave information completeness 60% of weight, fluency of delivery 30% and target language quality 10%. This scheme was more or less similar to the previous schemes reported in the literature that usually assigns a greater weight to information completeness and smaller weights to other criteria (see Han, 2016; Lee, 2008; Wang et al., 2015), although Collados Aís, Iglesias Fernández, Pradas Macías and Stévaux (2011) find that scores for information completeness are often skewed by fluency and target language quality.

Regarding the format of the rating scale, the author chose an eight-point rating scale, which can also be reduced to a four-band scale by collapsing two neighbouring points into a single score band (see Table 1). A similar scale structure has been used in interpretation assessment (see Han, 2015, 2016). Emerging evidence (e.g., monotonic development of Rasch-Andrich thresholds in line with the rating categories based on Rasch analysis) supported that rating scales of such kind could discriminate different performance levels reliably (see Han, 2015). It needs to be pointed out that regarding the dimension of *information completeness*, such descriptors as “80%” and “60-70%” were used as a heuristic guide to help raters classify students' performances into different categories or scale bands. It can be argued that based on these descriptors a rater may be unable to assign a band to an interpretation which renders 74% of original message in a source-language speech. However, in scoring *information completeness*,

Conceptualizing and Operationalizing a Formative Assessment Model

it is usually difficult for raters to pinpoint the exact percentage of interpretation correctly rendered. The issue at hand concerns the precision of measurement. Accordingly, the scalar descriptors were designed to accommodate this reality by asking raters to provide estimations rather than a precise percentage. For example, when an interpretation was found to have reproduced 73% of the original message (even though this exact percentage may be difficult for human raters to estimate), it would be appropriate to assign a band of 3 to the interpretation (i.e., 60-70%), and give it a band score of 6 (instead of 5).

When it comes to writing scalar descriptors, the author primarily drew on previous literature, especially Han (2015, 2016). A drafted version of the scale was then sent to the students to elicit their opinions. In week 3, the students provided their feedback. Although the students found the majority of the scalar descriptors acceptable, they objected to the percentage of source-language information that needed to be translated in each scale band. For example, in the draft version of the *information completeness* scale, the students needed to correctly translate at least 90% of the original message to obtain a score of 7 or 8. They argued that the standards were too stringent for undergraduate students, and needed to be dialled down. A compromise was made to reduce the percentage in each scale band by 10%. The rating scale was then finalized and distributed to all students (for the final version of the scale, see Table 1).

When evaluating an interpretation, all stakeholders needed to provide three pieces of information as formative feedback. Firstly, they needed to propose a band score for each of the three quality dimensions. Secondly, they should identify relevant evidence to justify why a given score was awarded. Thirdly, they needed to provide practical suggestions, based on the strengths and weaknesses identified, to guide students' future practice. Two major sections were therefore added to the rating scale, as can be seen in Table 1: the *Evidence & Justification* section where the stakeholders provide evidential justification for their proposed scores, and the *Diagnostics & Reflection* section where the stakeholders provide diagnostic analysis and offer concrete suggestions.

Table 1. Descriptor-referenced, scale-based assessment sheet

Student ID: Material ID: Time (Year-Month-Day):			
Band/Scoring Criteria	Information Completeness	Fluency of Delivery	Target Language Quality
Band 4 (Score range: 7-8)	A substantial amount of original messages delivered (i.e., > 80%), with a few cases of deviations, inaccuracies, and minor/major omissions.	Delivery on the whole fluent, containing a few disfluencies such as (un)filled pauses, long silence, fillers and/or excessive repairs.	Target language idiomatic and on the whole correct, with only a few instances of unnatural expressions and grammatical errors.
Band 3 (Score range: 5-6)	Majority of original messages delivered (i.e., 60-70%), with a small number of deviations, inaccuracies, and minor/major omissions.	Delivery on the whole generally fluent, containing a small number of disfluencies.	Target language generally idiomatic and on the whole mostly correct, with a small amount of instances of unnatural expressions and grammatical errors.
Band 2 (Score range: 3-4)	About half of original messages delivered (i.e., 40-50%), with many instances of deviations, inaccuracies, and minor/major omissions.	Delivery rather fluent. Acceptable, but with regular disfluencies.	Target language to a certain degree both idiomatic and correct. Acceptable, but contains many instances of unnatural expressions and grammatical errors.
Band 1 (Score range: 1-2)	A small portion of original messages delivered (i.e., < 30%), with frequent occurrences of deviations, inaccuracies, and minor/major omissions, to such a degree that listeners may doubt the integrity of renditions.	Delivery lacks fluency. It is frequently hampered by disfluencies, to such a degree that they may impede comprehension.	Target language stilted, lacking in idiomaticity, and containing frequent grammatical errors, to such a degree that it may impede comprehension.
Proposed score			
Evidence & Justification			
Diagnostics & Reflection			

Assessment Training

In order to make the rating scale and assessment procedure explicit to the students, the author organized three one-hour training sessions after class in which all students participated. Firstly, the author provided each student with an electronic copy of the assessment sheet and asked them to familiarize with its structure and content. Secondly, the author provided detailed explanations to key descriptors such as deviation, omission and disfluency features, each of which was also illustrated with written and/or audio samples. The students were encouraged to ask questions if further clarification was needed. Thirdly, the author explained how the rating scale could be used in assessing interpretation. Particularly, the students were asked to pay attention to critical features of a given interpretation, pinpoint relevant evidence to match scalar descriptors, and categorize the interpretation into one of the four scale bands before assigning it with a band score (i.e., one to eight). The author also instructed the students to assess *information completeness* first and then move to the other criteria (for a different procedure, see Yeh & Liu, 2006). The decision to rate *information completeness* first is based on three reasons. The first reason is that fidelity or *information completeness* has long been upheld as a predominant property of interpretation (e.g., Gile, 1999; Strong & Rudser, 1985; Llewellyn-Jones, 1981), a view that is also supported by surveys of interpreters (e.g., Bühler, 1986) and interpreter trainers (Lee, 2008; Lee, 2015). The second reason is that a similar rating strategy is recommended by Wu (2010), based on his findings from a sample of 30 raters assessing students' audio-recorded simultaneous interpretations. The third reason is that previous studies produced some findings that suggest the usefulness of the rating sequence (see Han, 2017). The next step in the training session is that the author provided to the students with a batch of 15 recordings that were produced in previous semesters for pilot rating. The students were also asked to compare scores with one another and justify why a specific score was given. Lastly, an assessment sheet filled by previous students was used as an exemplar to help the students appreciate how qualitative feedback could be provided.

Self and Peer Assessment

After an assessment was conducted, the author saved and collected interpretation recordings via the master console of the interpreting teaching system, and distributed recordings to each student for self-assessment. In addition, the author randomly selected interpretation recordings of four to six students (a total of eight to 12 recordings), and sent them to a random student for peer assessment. A student's performance was therefore assessed by four to six peers, using the performance assessment sheet. The students needed to complete one assessment sheet for each recording. Filled sheets needed to be returned to the author via email within three to four days after the assessment. The filled sheets were then checked and reviewed by the author, before peer assessment results were sent to each student.

Teacher Assessment

After each assessment, the author organized a panel of six raters who were familiar with the assessment scheme to evaluate the students' interpretation recordings. Two of the raters were university lecturers of interpreting (including the author), and the other four were teaching assistants who had gained rating experience in the previous semester. The raters gathered together, but worked independently, using the rating scale to assign numeric scores to the students' recordings. After each assessment, the author

sent each student a score report, which included self-ratings (derived from self-assessment), averaged peer ratings and averaged teacher ratings on each scoring dimension for each interpreting direction. The average score from the six raters represented a student's final score. Based on raters' average scores, the students (identified by their student number) were ranked, which were then made public to the whole class. Publicizing the scores may be an unconventional practice in many places. However, in the Chinese educational context, students' final grades are sometimes made public, although this practice is becoming increasingly unpopular in recent years.

Apart from the quantitative information, the author provided qualitative feedback to each student (the other five raters were not asked to do so; they only provided numeric scores). In general, the author provided two forms of feedback: individualized feedback and collective feedback.

In providing individualized feedback, the author relied on the performance assessment sheet to comment on each quality dimension of interpreting performance and gave suggestions for future practice. Particularly, to diversify feedback practice, the author alternated between written and audio feedback. While written feedback is a common practice in interpreter training (e.g., Choi, 2006; Lee, 2005; Postigo Pinazo, 2008), it seems that audio feedback is less used (see Witter-Merithew et al., 2001). In audio feedback a teacher records her/his appreciations, critiques, suggestions and other types of evaluative comment in an audio file that will then be sent to students. Arguably, compared to written feedback, audio feedback is more informative, engaging, and efficient (e.g., Lunt & Curran, 2010). Specifically, after the first assessment in week 5 the author sent each student written feedback in a Word document (usually 150 – 200 words); after the second assessment in week 10 the author gave them audio feedback in a MP3 file (usually 6 – 10 minutes); for the last assessment in week 15 the author switched back to written feedback. To provide prompt feedback to each individual, the author started evaluating students' recordings on the same day when formative assessment took place, and returned feedback immediately after a student's performance was scored and commented on. In other words, the author did not wait for all recordings to be assessed before sending comments back to the students. Instead, the author provided feedback to individual students on the go. In doing so, almost half of the students could receive feedback one day after an assessment, although the whole evaluation process usually took three days. It is worth noting that the order of students who received feedback first changed from one occasion to the next. One may argue that given the relatively large number of students involved in the assessment, teacher feedback practice can be adjusted when all assessments are over. However, the primary aim here was to generate sufficient and prompt feedback to each student, as soon as strengths and weaknesses of each student's performance were identified by the teacher.

After all recordings were assessed, the author summarized and listed common problems in the students' interpretation. Then the author discussed the problems in the next class to raise awareness, provide general solutions and suggest remedial measures. Such feedback provided to the whole class is what the author called collective feedback.

To sum up, in each assessment each student evaluated and reflected on her/his interpretation, obtained peer feedback from four to six classmates, and received individualized and collective feedback from the author/teacher.

Survey of Student Satisfaction

By the end of the 7th semester (i.e., the first semester of the fourth year) when all formative assessments were completed, the author designed a questionnaire to elicit students' opinions on how they perceived the assessment practice. Specifically, the author designed a questionnaire, using *SurveyMonkey*, which consisted of seven sections. In Section I, the students needed to provide their student ID numbers. The author did not intend to identify students based on their ID, as the students were reassured in the questionnaire that they could air their opinions freely. However, in hindsight this design may affect students' answers, and is therefore acknowledged in the discussion section as a potential limitation. Sections II to VI included questions clustered under five themes: 1) the characteristics of the formative assessments; 2) assessment criteria and rating scale; 3) self, peer and teacher assessment; 4) formative feedback, and 5) assessment results. Section VII was a general question, asking the students to rate their overall satisfaction level regarding the formative assessment model. Although the students had some prior experience of self-assessment, such an elaborate formative assessment practice as described above was new to them. It was therefore assumed that when responding to the questionnaire the students may largely refer to how they perceived the current assessment system as it was.

For Sections II to VI, a total of 39 statements were created. To ascertain to what extent the students agreed with each statement, a seven-point rating scale was used, assigning a number to each descriptor (e.g., "Strongly agree" – 7, "Neutral" – 4, "Strongly disagree" – 1). In addition, three statements were reversed (statements 6 and 11 in Section II and statement 6 in Section V), which needed to be adjusted when calculating final ratings. For a view of the 39 statements in the questionnaire, please see Tables 2, 3, 4, 5 and 6. It should be pointed out that the online questionnaire was originally developed in Chinese, and completed by all students ($n = 41$). The questionnaire statements presented in the chapter are an English translation of the Chinese version.

PRELIMINARY EVALUATION OF THE FORMATIVE ASSESSMENT MODEL

This section presents results of preliminary evaluation of the formative assessment model from two types of stakeholders: the students and the teacher (the author). The students evaluated the model and shared their thoughts primarily via the final questionnaire, whereas the teacher's evaluation was based on his experience of and reflection on operationalizing the model.

Students' Satisfaction: Results From the Questionnaire

In the following section, the questionnaire results are presented and organized according to the six major sections in the questionnaire (i.e., Sections II to VII).

Characteristics of the Formative Assessment

Table 2 presents the averaged ratings for the 11 statements in Section II of the questionnaire. In general, the students thought positively of the assessments and the materials used in the assessments, as most statements scored above five on the seven-point scale. It is worth noting that statement three (i.e., II-03)

Table 2. Questionnaire results for Section II

Statements	Averaged Ratings
II-01. The number (frequency) of formative assessment conducted in this semester was adequate.	5.49
II-02. The interval between the assessments was appropriate.	5.56
II-03. It was a good practice to provide background information and relevant vocabulary before the assessment.	6.49
II-04. The time when background information and relevant vocabulary was provided was acceptable.	5.85
II-05. The number of students (about ten students) who took the assessment together in the language lab was appropriate.	6.12
II-06. I was not distracted by other students during the assessment.	3.49
II-07. The number of interpreting tasks in each assessment was appropriate.	5.80
II-08. The duration of source-language speeches was appropriate.	5.88
II-09. The audio files of source-language speeches were of high quality.	6.39
II-10. The topical content of source-language speeches was diverse.	5.51
II-11. The source-language texts were too difficult.	4.27

scored highest with an average rating of 6.49, which suggests that sending background information to the students prior to assessment was positively perceived. This may be because gaining general knowledge of interpreting topics and doing some preparation work would ease students' test anxiety and potentially improve their performance (see Díaz-Galaz, Padilla, & Bajo, 2015). Another interesting observation is that although the students felt comfortable with the number of students taking the assessment together (II-05), they did express their concern of being distracted by their peers (II-06). One way to explain this result is that while the students wore sound-proof headphones when interpreting, the open-plan language lab did not provide private space for each student. As a result, some students may be sensitive to and distracted by others. An additional noteworthy finding is that the students as a whole rated statement 11 neutrally (II-11). In other words, some students believed that the source-language texts were difficult, while others did not. Although assessment tasks generally need to differentiate test takers, especially in proficiency testing, in the formative assessment context using tasks that are too difficult for some students while too easy for others is less likely to maximize educational benefits. Caution is therefore advised when designing assessment tasks for the purpose of formative assessment.

Assessment Criteria and Rating Scale

Table 3 shows the averaged ratings for the five statements in Section III of the questionnaire. As can be seen in the table, the students rated statements one to four relatively high, with the ratings hovering around 5.5. This suggests that the students generally supported and understood the assessment criteria, and believed that they were adequately trained. However, they were less confident in applying the rating scale to assess interpretation, as indicated by statement five (III-05). This seemingly contradictory result may be due to the fact that the students lacked practice in using the rating scale prior to formal self and peer assessment, although they had gained much conceptual understanding. It may also be because in applying the rating scale the students encountered situations in which they felt uncertain and difficult to decide on a rating. These experiences then influenced their opinions on statement five.

Conceptualizing and Operationalizing a Formative Assessment Model

Table 3. Questionnaire results for Section III

Statements	Averaged Ratings
III-01. The assessment criteria were reasonable, given the current stage of my learning.	5.61
III-02. The weighting scheme of the assessment criteria was appropriate.	5.54
III-03. The scalar descriptors for each assessment criterion were clear.	5.66
III-04. The teacher provided sufficient training and guidance regarding how to use the rating scale.	5.49
III-05. I was confident in using the rating scale to assess interpretation quality.	4.85

Self, Peer, and Teacher Assessment

Table 4 summarizes the averaged ratings for the seven statements in Section IV of the questionnaire. Based on the table, two observations could be made. The first observation is that overall the students thought highly of the number of assessors involved in peer and teacher assessment (IV-05 and IV-07). The second observation is that the students perceived self and peer assessment as being positive and beneficial (IV-01 and IV-03), but they rated the reliability and fairness of self-ratings and peer ratings relatively low (see IV-02 and IV-04). By contrast, teacher-generated ratings were perceived by the students to be much more reliable and fairer, with the rating above 6.00 (see IV-06). This result suggests that the perceived benefit of formative assessment may not lie in the multiple ratings generated by students themselves and their peers, but in the assessment practice itself, which gives students opportunities to reflect on their performance and learn from others. In addition, in the formative assessment context in which self, peer and teacher assessment is instituted as paralleled practices, teachers' ratings are still considered to be more reliable and authoritative.

Formative Feedback

Table 5 presents the averaged ratings for 11 statements in Section V of the questionnaire. It seems that both peer and teacher feedback were found helpful by the students (see V-01 and V-03), but teacher feedback was conceived as being more transparent and detailed than peer feedback (see V-05 and V-02). Nevertheless, the students did not regard the scope of teacher feedback as being comprehensive enough (V-06). This result highlights a dilemma faced by the teacher. Given that the available time to

Table 4. Questionnaire results for Section IV

Statements	Averaged Ratings
IV-01. Self-assessment was beneficial to me.	6.00
IV-02. Self-ratings were reliable and fair.	4.63
IV-03. Peer assessment was beneficial to me.	5.85
IV-04. Peer ratings were reliable and fair.	4.85
IV-05. The number of peers (four to six peers) who assessed my interpretations was sufficient.	5.80
IV-06. Teacher-generated ratings were reliable and fair.	6.10
IV-07. The number of raters (two teachers and four teaching assistants) involved in assessment was satisfactory.	5.83

Table 5. Questionnaire results from Section V

Statements	Averaged Ratings
V-01. Peer feedback was helpful to me.	5.76
V-02. The content of peer feedback was transparent and detailed.	4.83
V-03. Teacher feedback was helpful to my interpreting study.	6.39
V-04. The formats of teacher feedback (i.e., written and audio feedback) were actionable.	6.44
V-05. The content of teacher feedback was transparent and detailed.	6.27
V-06. The content of teacher feedback was not comprehensive.	4.68
V-07. The feedback I received was prompt.	5.90
V-08. I was able to practice interpreting in my spare time and to work on my weaknesses, based on the feedback.	4.56
V-09. The feedback inspired me to practice interpreting in my spare time.	5.22
V-10. I would read or listen to the feedback repeatedly so as to understand my weaknesses in interpreting.	4.73
V-11. I would formulate my learning plan based on the feedback.	4.71

assess and comment on students' interpretations was very limited (as the author also wanted to provide prompt feedback), it would be difficult, if not impossible, for the author to comment on every aspect of a student's performance in a detailed fashion. On a positive note, the students generally gave high ratings to the promptness of feedback (V-07) and perceived positively of the different formats of feedback (V-04). Regarding the use of teacher feedback (i.e., feedback from the author as a teacher/evaluator), the results were not very optimistic (see V-08, V-09, V-10 and V-11), with three of the four ratings being below 5.00. Considering that the students thought highly of the feedback, it came as a surprise that they did not make best use of the feedback. Perhaps, more guidance is needed to encourage the students to effectively use feedback in future.

Assessment Results

Table 6 displays the averaged ratings for five statements in Section VI of the questionnaire. As can be seen in the table, it would appear that the students were generally supportive of how their assessment results were reported. However, it can also be seen that the averaged ratings for statements two and three (VI-02 and VI-03) were slightly lower than those of the other statements. This pattern suggests that compared to the other practices of reporting assessment results the students were less favourable of publicizing their final scores.

Table 6. Questionnaire results for Section VI

Statements	Averaged Ratings
VI-01. The score report I received was detailed and clear.	5.68
VI-02. It was appropriate to make public the final scores after each assessment.	5.22
VI-03. It was acceptable to rank students according to their scores.	5.41
VI-04. It was reasonable to report scores separately for each interpreting direction.	6.17
VI-05. I found it satisfactory to use the average score (from the six raters) as my final score.	5.61

Overall Satisfaction Level

Table 7 shows a detailed picture of how the students rated the formative assessment model as a whole. In general, the students seemed to be satisfied with the formative assessment experience, with the averaged rating of almost 6.00. The break-down analysis, however, is a bit concerning. Granted, 23 students (about 56%) endorsed the category of “Satisfied”, and 10 students (about 25%) chose “Very satisfied”. But one student thought very negatively of the assessment practices, as s/he endorsed “Very dissatisfied”. In order to find out why this particular student was dissatisfied, the author conducted a brief follow-up interview. It turned out that the student simply did not like any kind of testing and assessment, particularly assigning scores to her/himself and being assigned scores by her/his peers and teachers. The frequent assessments made her/him feel uncomfortable and jittery. Similar finding is also reported by Evans, McKenna and Oliver (2005) in which students dislike assessing themselves or their peers, and would prefer to have teachers assess them.

Teacher’s Reflections: Good Practices and Potential Problems

By implementing the assessment model and by reading the students’ questionnaire results, the author gained a better understanding of how formative assessment could be operationalized to benefit students’ learning. The author would like to highlight some practices that were found to be beneficial to the students and conducive to implementation of formative assessment. One of the good practices is to inform students of why formative assessment needs to be conducted, and why self and peer assessment are important components of formative assessment. In other words, teachers need to justify the practice of formative assessment. Otherwise, some students may become uncooperative, questioning the value of formative assessment and even seeing self and peer assessment as teachers’ pretext to avoid grading assignments (Boud, 1992). Another good practice is to explicate assessment criteria before assessment (Schjoldager, 1996; William & Thompson, 2007), preferably getting students involved in conceptualizing and devising criteria (also see Lee, 2016; Orsmond et al., 2000). This enables students to gain an in-depth understanding of the standards against which their performance is evaluated. In addition, teacher’s engagement with students at this stage empowers students and makes them accountable for their future learning. An additional good practice is to provide sufficient guidance to students on using assessment criteria to assess and critique interpretation, especially for those students who have no prior experience of self and/or peer assessment. Previous literature suggests that self-assessing or assessing other’s work is not an easy task for some students (Boud, 1992; Fowler, 2007; Ibrahim-González & Noordin, 2012). They may find it difficult to evaluate their own performance accurately or feel reluctant to pass judgement to others’. As can also be seen from the questionnaire results, the students were

Table 7. Results for the overall satisfaction level

To what extent are you satisfactory with the formative assessments?							
Very Dissatisfied (1)	Dissatisfied (2)	Somewhat Dissatisfied (3)	Neutral (4)	Somewhat Satisfied (5)	Satisfied (6)	Very Satisfied (7)	Averaged Rating
1 (2.44%)	0 (0%)	0 (0%)	1 (2.44%)	6 (14.63%)	23 (56.1%)	10 (24.39%)	5.93

not very confident in using the rating scale to assess interpretation, even though they received training and they reported that they understood assessment criteria well. The result suggests that teachers need to provide more practical guidance to students on applying assessment criteria. More importantly, in order to generate useful feedback, guidance also needs to be provided to help students become providers of useful feedback (Sadler, 1998, p. 78). Essentially, these good practices boil down to establishing a trustworthy relationship between teachers and students, getting students involved and informed, and providing sufficient support before, during and after assessment.

Meantime, the author also identified several potential problems that could undermine the effectiveness of formative assessment. One concerning problem is that students may be inactive and unresponsive to feedback. As the questionnaire results show (see Section V of the questionnaire), although the students appreciated peer and teacher feedback, it seems that the feedback had not inspired all of them to take concrete actions after the assessments (e.g., doing more practice, refining certain skills, addressing weaknesses). In other words, some students did not use the feedback effectively. Granted, one of the most important goals of formative assessment is to generate informative feedback to students so that they could engage with and utilize feedback to benefit their learning. But if students lack motivation to capitalize on feedback, it would defeat the very purpose of formative assessment. To help students realize the importance of feedback and act upon feedback, there is a need to ensure that self, peer and teacher feedback are related to one another. In the current practice, it would seem that self, peer and teacher feedback represent different ways to improve interpreting performance. In some cases, if contradictory comments or suggestions are found in self, peer and teacher feedback, it would send confusing messages to students, thus detracting from the value of feedback. Therefore, these three kinds of feedback should be related and coordinated in a principled manner to maximize their benefits. Another potential problem relates to the lack of carefully designed assessment tasks that can be used in formative assessment. In the current study, three interpreting tasks were created for each direction, which represented a very small pool of tasks. In hindsight, the author felt a genuine need to develop a *task bank* consisting of a large quantity of tasks that are developed by experienced teachers based on a set of task specifications, and that are grouped under different topics and difficulty levels. These tasks can be carefully sampled for formative assessments, and selectively distributed to students of different ability levels as enhancement assignments outside of the class. The availability of and accessibility to a task bank would contribute greatly to formative assessment. An additional problem centers on the practice to publicize students' assessment results after each assessment. As suggested by the questionnaire results (Section VI), not all students were happy about letting others know their scores. This reaction is understandable, given that some students may regard assessment results as a very private matter. Publicizing students' final scores may therefore weaken their motivation to participate in future formative assessment. Finally, there is a potential limitation of the questionnaire design in that the students were asked to report their ID numbers in the questionnaire. Although the author/teacher could not directly identify who a student is based solely on their ID, there is a distinct danger of exposing students' identity, which could undermine the validity of their answers to questionnaire items. As such, asking for student ID numbers or other personal information in the questionnaire should be avoided in future practice.

Apart from the good practices and the potential problems, the author also wants to foreground a main drawback of the formative assessment model: the substantially increased workload for the author/teacher. Usually, the author needed to fulfil at least five major duties: 1) managing and coordinating the assessment practice and serving as a point of contact, 2) reviewing self and peer assessments as well as organizing, compiling and distributing peer feedback to each student, 3) evaluating a total of 81

Conceptualizing and Operationalizing a Formative Assessment Model

three-minute interpretation recordings (i.e., 41 students × 2 directions) each time, 4) providing feedback (written or verbal) to each student based on their performance, and 5) collecting and organizing scores from the other five raters and preparing a final score report to the students. Fulfilling these duties was labor-intensive and time-consuming. To some extent, the heavy workload renders the formative assessment model less practical and actionable.

An additional interesting topic relates to whether there was a difference between the self, peer, and teacher evaluation in terms of the scoring, although a detailed analysis and description is beyond the scope of the chapter (for details, see Han 2018). According to the questionnaire results, the students thought that the teacher ratings were more relevant and represented a fairer measure of their performance, probably because they found that they differed from teacher assessments in absolute terms. Indeed, when the teachers' ratings were taken as a yardstick, the students as a group tended to overestimate not only their own performances, but also their peers'. In addition, such overestimation was found to be statistically significant (see Han 2018). However, correlational analyses showed that the strength of association between self and teacher assessments as well as between peer and teacher assessments was fairly strong overall, indicating that the students were able to rank-order interpreting performances in a similar fashion as the teachers did.

FUTURE DIRECTIONS

The current model of formative assessment has much room for improvement. Firstly, customized learning platforms such as *Moodle*, *Dashboard* and *Google Classroom* could be used to provide online delivery of assessment-related materials, facilitate submission of self and peer assessment results, optimize teacher's feedback practice, and ultimately reduce teacher's workload. Secondly, in addition to self and peer assessment, teachers could incorporate other formative assessment practices such as group/colaborative assessment, portfolio assessment and dynamic assessment. The goal is to identify an optimal combination of assessment practices that suits the needs of a particular group of students and that could best promote student learning in a certain context. Thirdly, there is a need of task banking for formative assessment. Interpreting tasks could be developed based on detailed task specifications to represent different developmental stages of interpreting ability. Fourthly, it is advisable to provide an individual progress record for each student so that they could track and review their study, and that teachers could better evaluate the effectiveness of formative assessment. Fifthly, teachers could give more attention to students' actual use of feedback. That is, there is a need to monitor how students engage with feedback and to investigate the effects of feedback on students.

CONCLUSION

Formative assessment has been increasingly valued by interpreter trainers and educators, as it has the potential to promote student learning, foster reflective practices and nurture autonomous learners. This chapter describes the design and implementation of a formative assessment model for an undergraduate-level interpreting course, and reports on the students' evaluation of and teacher's reflection on the assessment model. Although some good practices emerge from the assessment model (e.g., involving students in devising assessment criteria, providing rater training), a number of potential problems are

also identified. Particularly, the inefficient use of feedback by the students, the lack of a task bank and the heavy workload for the teacher are some of the major issues to be addressed. Going forward, interpreter trainers could avail themselves of online learning platforms to improve work efficiency, build task banks to facilitate flexible formative assessment, and explore ways to monitor students' engagement with feedback.

ACKNOWLEDGMENT

The author would like to thank the anonymous reviewers and book editors for their valuable and constructive feedback on the earlier version of this chapter. This work was supported by the Fundamental Research Funds for the Central Universities [SWU1709617] and by the China Postdoctoral Science Foundation [2017M622938].

REFERENCES

- Arumí Ribas, M. (2010). Formative assessment in the interpreting classroom: Using the portfolio with students beginning simultaneous interpreting. *Current Trends in Translation Teaching and Learning E*. Retrieved from <http://www.cctl.org/uploads/5/2/4/3/5243866/currenttrendstranslation2010.pdf>
- Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education*, 85(5), 536–553. doi:10.1002/ce.1022
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. doi:10.1080/0969595980050102
- Bloom, B. S., Hastings, J. T., & Madaus, G. F. (Eds.). (1971). *Handbook on the formative and evaluation of student learning*. New York: McGraw-Hill.
- Boud, D. (1992). The use of self-assessment schedules in negotiated learning. *Studies in Higher Education*, 17(2), 185–200. doi:10.1080/03075079212331382657
- Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.
- Choi, J. Y. (2006). Metacognitive evaluation method in consecutive interpretation for novice learners. *Meta*, 51(2), 273–283. doi:10.7202/013256ar
- Collados Aís, Á., Iglesias Fernández, E., Pradas Macías, E. M., & Stévaux, E. (2011). *Qualitätsparameter beim Simultandolmetschen: Interdisziplinäre Perspektiven*. Tübingen: Gunter Narr.
- Díaz-Galaz, S., Padilla, P., & Bajo, M. T. (2015). The role of advance preparation in simultaneous interpreting: A comparison of professional interpreters and interpreting students. *Interpreting*, 17(1), 1–25. doi:10.1075/intp.17.1.01dia
- Evans, A. W., McKenna, C., & Oliver, M. (2005). Trainees' perspectives on the assessment and self-assessment of surgical skills. *Assessment & Evaluation in Higher Education*, 30(2), 163–174. doi:10.1080/0260293042000264253

Conceptualizing and Operationalizing a Formative Assessment Model

- Fowler, Y. (2007). Formative assessment: Using peer and self-assessment in interpreter training. In *The critical link 4: Professionalisation of interpreting in the community* (pp. 253-262). Amsterdam: John Benjamins.
- Gile, D. (1999). Variability in the perception of fidelity in simultaneous interpretation. *Hermes*, 22, 51–79.
- Han, C. (2015). Investigating rater severity/leniency in interpreter performance testing: A multifaceted Rasch measurement approach. *Interpreting*, 17(2), 255–283. doi:10.1075/intp.17.2.05han
- Han, C. (2016). Investigating score dependability in English/Chinese interpreter certification performance testing: A generalizability theory approach. *Language Assessment Quarterly*, 13(3), 186–201. doi:10.1080/15434303.2016.1211132
- Han, C. (2018). *A longitudinal quantitative investigation into the validity of self and peer assessment applied to English-Chinese bi-directional interpretation in an undergraduate interpreting course*. Manuscript submitted for publication.
- Han, C. (2017). Using analytic rating scales to assess English-Chinese bi-directional interpretation: A longitudinal Rasch analysis of scale utility and rater behavior. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 16.
- Han, C., & Riazi, M. (2017). Investigating the effects of speech rate and accent on simultaneous interpretation: A mixed-methods approach. *Across Languages and Cultures*, 18(2), 237–259. doi:10.1556/084.2017.18.2.4
- Han, C., & Riazi, M. (2018). The accuracy of student self-assessments of English-Chinese bidirectional interpretation: A longitudinal quantitative study. *Assessment & Evaluation in Higher Education*, 1–13. doi:10.1080/02602938.2017.1353062
- Han, C., & Slatyer, H. (2016). Test validation in interpreter certification performance testing: An argument-based approach. *Interpreting*, 18(2), 231–258. doi:10.1075/intp.18.2.04han
- Hartley, A., Mason, I., Peng, G., & Perez, I. (2003). *Peer- and self-assessment in conference interpreter training*. Retrieved from <https://www.llas.ac.uk/resourcedownloads/1454/hartley.rtf>
- Iaroslavschi, M. (2011). Becoming an interpreter: Assessment and self-assessment in untutored practice sessions. *Cognition, Brain, Behavior: An Interdisciplinary Journal*, 15(2), 229–249.
- Ibrahim-González, N., & Noordin, N. (2012). Learner autonomy via self-assessment in consecutive interpreting for novice learners in a non-interpreting environment. *Sino-US English Teaching*, 9(4), 1065–1073.
- Lee, J. (2008). Rating scales for interpreting performance assessment. *The Interpreter and Translator Trainer*, 2(2), 165–184. doi:10.1080/1750399X.2008.10798772
- Lee, S.-B. (2016). University students' experience of 'scale-referenced' peer assessment for a consecutive interpreting examination. *Assessment & Evaluation in Higher Education*, 42(7), 1015–1029. doi:10.1080/02602938.2016.1223269

- Lee, Y.-H. (2005). Self-assessment as an autonomous learning tool in an interpretation classroom. *Meta*, 50(4). doi:10.7202/019869ar
- Lee, Y.-H. (2011). Comparing self-assessment and teacher's assessment in interpreter training. *T&I Review*, 1, 87-111.
- Lim, H.-O. (2013). A jury of your peers: Peer assessment in consecutive interpretation pedagogy. *Forum*, 11(1), 77-97. doi:10.1075/forum.11.1.04lim
- Llewellyn, J. P. (1981). Simultaneous interpreting. In J. K. Woll & M. Deuchar (Eds.), *Perspectives on British Sign Language and Deafness* (pp. 89-104). London: Croom Helm.
- Lunt, T., & Curran, J. (2010). 'Are you listening please?' The advantages of electronic audio feedback compared to written feedback. *Assessment & Evaluation in Higher Education*, 35(7), 759-769. doi:10.1080/02602930902977772
- Mead, P. (2005). Methodological issues in the study of interpreters' fluency. *The Interpreters' Newsletter*, 13, 39-63.
- Moser-Mercer, B. (1996). Quality in interpreting: Some methodological issues. *The Interpreters' Newsletter*, 7, 43-56.
- Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education*, 25(1), 23-38. doi:10.1080/02602930050025006
- Pöchhacker, F. (1993). On the science of interpretation. *The Interpreters' Newsletter*, 5, 52-59.
- Postigo Pinazo, E. (2008). Self-assessment in teaching interpreting. *TTR: Traduction, Terminologie. Redaction*, 21(1), 173-211.
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28(1), 4-13. doi:10.1002/bs.3830280103
- Russo, M. (1995). Self-evaluation: The awareness of one's own difficulties as a training tool for simultaneous interpretation. *The Interpreters' Newsletter*, 6, 75-85.
- Russo, M. (2011). Aptitude testing over the years. *Interpreting*, 13(1), 5-30. doi:10.1075/intp.13.1.02rus
- Sadler, D. R. (1998). Formative assessment: Revising the territory. *Assessment in Education: Principles, Policy & Practice*, 5(1), 77-84. doi:10.1080/0969595980050104
- Sandrelli, A. (2015). Becoming an interpreter: The role of computer technology. *MonTI*, 2, 111-138.
- Sawyer, D. B. (2004). *Fundamental aspects of interpreter education: Curriculum and assessment*. Amsterdam: John Benjamins. doi:10.1075/btl.47
- Schjoldager, A. (1996). Assessment of simultaneous interpreting. In C. Dollerup & V. Appel (Eds.), *Teaching Translation and Interpreting 3: New Horizons* (pp. 187-195). Amsterdam: John Benjamins. doi:10.1075/btl.16.26sch

Conceptualizing and Operationalizing a Formative Assessment Model

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (Vol. I, pp. 39–83). Chicago, IL: Rand McNally.

Strong, M., & Rudser, S. F. (1985). An assessment instrument for sign language interpreters. *Sign Language Studies*, 49(1), 343–362. doi:10.1353/ls.1985.0008

Wang, J.-H., Jemina, N., Goswell, D., & Carmichael, A. (2015). The design and application of rubrics to assess signed language interpreting performance. *The Interpreter and Translator Trainer*, 9(1), 83–103. doi:10.1080/1750399X.2015.1009261

William, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Mahwah, NJ: Erlbaum.

Witter-Merithew, A., Taylor, M., & Johnson, L. (2001). Guided self-assessment and professional development planning: A model applied to interpreters in educational settings. In C. Nettles (Ed.), *Proceedings of the 17th national conference of the registry of interpreters for the deaf* (pp. 153-226). Alexandria, VA: RID Publications.

Wu, S. C. (2010). *Assessing simultaneous interpreting: A study on test reliability and examiners' assessment behavior* (Unpublished doctoral dissertation). Newcastle University, UK.

Yeh, S.-P., & Liu, M. (2006). 口譯評分客觀化初探:採用量表的可能性 [A more objective approach to interpretation evaluation: Exploring the use of scoring rubrics]. *Journal of the National Institute for Compilation and Translation*, 34(4), 57–78.

ENDNOTE

¹ Bachelor degree programmes in mainland China take four years (usually eight semesters in total).

Chapter 5

Competency–Based Education and Assessment: A Proposal for US Court Interpreter Certification

Melissa Wallace

University of Texas at San Antonio, USA

ABSTRACT

In an attempt to analyze the reliability and validity of the most frequently used oral certification exams for court interpreters in the United States, this chapter examines the basic test model used for state-level certification through the lens of concepts in testing theory. Having identified several limitations to the currently used performance-based model, a hybrid model which includes competency-based education and assessment is proposed. By building on best practices in competency-based education, the alternative credentialing paradigm proposed here would represent an innovation in the context of court interpreter certification in the United States, requiring the transfer of assessment criteria usually used in traditional educational contexts into the realm of professional training. The proposed hybrid model would necessitate a shift from one high-stakes exam to assessment of a series of compartmentalized competency clusters that would account for soft skills and dispositional traits not currently assessed.

INTRODUCTION AND OVERVIEW

In the face of doubt about whether or not current training experiences adequately prepare interpreters to enter the field, assessment and accreditation¹ practices are necessarily also placed under the spotlight. This is especially true in the case of the state-level court interpreting oral exams (referred to frequently as the “Consortium” exams) used in the United States. As the most frequently-used gateway credential to certified court interpreting in the nation, notions of quality in assessment (and indirectly, in training for the exams) are of vital importance and merit close examination².

DOI: 10.4018/978-1-5225-5225-3.ch005

Competency-Based Education and Assessment

This chapter takes to task the quality of the construction of the Consortium exams, analyzing them through the lens of testing theory. Upon identifying several limitations to the current model, competency-based education and assessment offer a promising remedy to fill the gap in current training offerings. Furthermore, having identified a series of knowledge, skills and abilities (KSAs), dispositional traits and soft skills which are considered crucial for court interpreters but which are not currently assessed, this chapter proposes a new, hybrid model for court interpreter certification that combines competency-based *and* performance-based testing in order to improve the validity and reliability of current credentialing practices.

Barriers to Reliable Interpreter Assessment

Interpreting studies stakeholders, from practicing professionals to researchers, express broad consensus on the significance of and justification for further inroads into empirical studies focusing on assessment and testing in interpreting studies (see Iglesias Fernández, 2013 and Niska, 2002). Assessment is vitally important not only for the purpose of screening applicants for entry into educational programs, providing feedback for students as they progress in their training, or testing their knowledge and skills at the end of a course of study, but most germane to the present discussion, it is also essential for qualifying exams such as the certification exams used in the field of court interpreting.

A burgeoning emphasis on quality assurance in the pan-European context has recently arisen as an essential element “for ensuring legal certainty and enhancing mutual trust between countries and their respective judicial systems” (Giambruno, 2016, p. 117) as recent legislation, particularly *Directive 2010/64/EU of the European Parliament and of the Council of 20 October 2010 on the Right to Interpretation and Translation in Criminal Proceedings*³ impacts the training, credentialing and register-building of the 28 Member States. In the American context, quality assurance is equally challenging. Nonetheless, providing competent interpreters in the justice system is not only a moral imperative: it is a constitutionally guaranteed protection.

Identifying and certifying competent court interpreters represents one of the cornerstones of due process in the US judicial system. Due process refers to the guarantee that individuals shall not be deprived of life, liberty, or property without notice and an opportunity to defend themselves. Language access is one of the foundations of due process, as to understand (and to be understood) by the courts is indispensable to full and active participation in one’s own defense. To be sure, the urgency of continuing to make inroads in interpreting studies in the realm of assessment bears most directly upon credentialing exams.

Importance notwithstanding, barriers to the provision of quality language services in the justice system are as varied as they are ubiquitous and well-documented. These can range from a lack of compliance with existing language access legislation (Wallace, 2015), difficulty in training, recruiting and credentialing qualified practitioners, the use of ad hoc or inadequately tested interpreters (Giambruno, 2016)⁴, a reluctance to pay for language services, outsourcing, and deficient accreditation schemes (Ozolins, 2016). Admittedly, other barriers to quality exist that fall fully outside the realm of education or testing, such as anti-immigrant sentiment or a lack of political will and administrative action (Giambruno, 2016). Some systemic challenges are directly specific to *administering* interpreting tests, including assembling a team of subject matter experts to design the constructs of the exams, building in protocols to assure rating consistency, or assuring equitable testing for candidates in all language combinations tested (see Skaaden & Wadensjö, 2014).

Gateways to the Profession: US Court Interpreter Certification

This section offers an overview of the two paths to certification currently available to aspiring court interpreters in the United States⁵. These two paths include the Federal Court Interpreter Certification Exam (available only in the Spanish – English language pair) and the state-level exams administered by the National Center for State Courts⁶. The latter, which include full or abbreviated oral exams for 18 language pairs (National Center for State Courts, 2017), are often referred to as state-level exams to distinguish them from the certification which allows interpreters to practice in federal courts. Most frequently, however, they are known as Consortium exams, in reference to the Consortium for Language Access in the Courts, the original collaborating entity of participating US states that pooled funds and resources in order to design, administer and rate certification exams. The state-level Consortium exams are discussed in more detail due to the fact that they are available in a greater combination of language pairs and are thus administered to the greatest number of examinees. It is important to bear in mind that both exams are performance-based and are summative assessment tools.

A brief history of court interpreter certification in the United States necessarily begins with the Federal Court Interpreter Certification Exam, or FCICE, administered under contract from the Administrative Office of U.S. Courts (AO) with the National Center for State Courts (NCSC), and created and implemented in direct response to the Court Interpreters Act of 1978. Signed by President Carter on October 29, 1978, the Court Interpreters Act established the right of any individual involved in federal proceedings to have a certified court interpreter if his/her communication or comprehension capabilities were inhibited because of a language barrier or a hearing or speech impairment. The resulting federal certification examination consists of a two-phase process involving a Spanish-English written examination (phase one) and an oral examination (phase two) administered on a biennial basis with the written exam and the oral exams taking place in alternating years. Interpreters must pass the written examination with a score of 75 percent or higher in order to be eligible to sit for the oral examination (de Jongh, 2012, p. 19). The written exam serves primarily as a screening test for linguistic competence in English and Spanish, and is a four-option, multiple choice test of job-relevant language ability. It tests comprehension of written text, knowledge of vocabulary, idioms, and grammatically correct expression, and the ability to select an appropriate target language rendering of source language text.

In contrast, the oral component of the federal exam directly measures *interpreting* skills. Because it fulfills the legal mandate for a criterion-referenced performance examination, the oral examination is the basis for certification to interpret in the federal courts, and consists of five parts: interpreting in the consecutive mode; interpreting a monologue in the simultaneous mode; interpreting a witness examination in the simultaneous mode; sight translation of a document from English into Spanish; and sight translation of a document from Spanish into English. To pass the federal oral examination, candidates must obtain an overall score of 80% or higher on all exercises (Paradigm Testing, 2017, p. 6). Although originally available for interpreters between English and Navajo, Haitian Creole and Spanish, the Spanish-English exams are the only language combination currently offered. In addition to the fact that federal certification is only available in one language combination, it is also characterized by notoriously low pass rates⁷.

Court certification at the state level through the National Center for State Courts, referred to as the “Consortium” exams in reference to the Consortium for Language Access in the Courts which was reconfigured into the Council of Language Access Coordinators in 2012⁸, is the credentialing system which has the most omnipresent role in the United States and is the axis around which the testing analyses in this chapter revolve. Certification by Consortium exam is the most commonly required credential in

Competency-Based Education and Assessment

the country as it allows court interpreters to practice in municipal, county and state venues and, most importantly, it is the only recognized certification for spoken languages to offer language combinations other than Spanish / English (National Center for State Courts, 2016). In fact, there are 18 language combinations currently available. Undoubtedly, the court interpreting certification exams administered by this entity hold absolute primacy as state-level exams are the most oft-used gateway to the court interpreting profession in the United States.

Consortium exams all adhere to the guidelines carefully delineated in the *Test Construction Manual*, the guidepost to oral exams for all language pairs offered. Consortium oral exams consist of four parts based on transcripts or other court documents which simulate actual court interpreting. Furthermore, and in large part thanks to the Consortium's commitment to abiding by *Test Construction Manual* (TCM) parameters (see National Center for State Courts, 2014), the Consortium has enjoyed consistency as well as a perception of face validity by interpreting stakeholders.

PERFORMANCE-BASED ACCREDITATION FOR US COURT INTERPRETERS

Certification exams identify candidates who are already able to meet a minimum standard in interpreting performance, acting as a threshold and thus allowing or impeding entry into the profession. Currently in the United States, court interpreter certification is entirely performance-based, in contrast to some countries in which interpreters are considered qualified to practice in court based solely on success on written exams or completion of specific undergraduate degrees⁹. The Federal Court Interpreter Certification Exam (FCICE) as well as the certification exams used at the state level (administered under the auspices of the National Center for State Courts) all require their candidates to pass exercises in simultaneous interpreting, consecutive interpreting, and sight translation (see also Feuerle, 2013). Performance-based assessments are those that ask examinees to perform the tasks that they would be expected to perform on the job.

Psychometric Considerations and Consortium Oral Exams

In the case of oral certification exams, whose purpose is to evaluate interpreter ability and not to gauge suitability for entry into a program of study, all valid and reliable assessments of skills depend upon the thoroughness with which testing instruments are psychometrically normed.

Psychometrics is the branch of psychology that “deals with the design, administration, and interpretation of quantitative tests for the measurement of psychological variables such as intelligence, aptitude, and personality traits” (ALTA Language Services, Inc., 2010, p. 33). Etilvia Arjona-Tseng is not alone in affirming that the need for a psychometric approach to aptitude testing in the realm of court interpreting is undisputed. In her discussion of the FCICE, she states that

[a] psychometric approach ... implies following a number of procedures to ensure construct validation, such as standardization of test administration procedures, proper pilot testing of instruments and adherence to standard principles of measurement and test construction. Rater training, for example, enhances reliability in the evaluation of candidates' performance. Perhaps equally important, a psychometrically sound approach enables us to evaluate whether we are, indeed, achieving our stated objective ... (1994, pp. 83-84).

In general, the evaluation of the psychometric features of a certification exam relate directly to their validity and reliability (Hlavac & Orlando, 2015)¹⁰. While this section does not purport to delve deeply into language testing theory, any discussion of quality in accreditation must necessarily examine some important methodological and psychometric properties of Consortium oral exams. Do they measure what they purport to measure, and are they adequate instruments in light of basic concepts in testing theory?¹¹ Cynthia Giambruno Miguélez states that

[i]n order to construct a valid and reliable certification instrument, it is necessary to understand the foundations of testing and the steps that are involved in the actual elaboration of the instrument. Many issues are involved in testing instrument design and development including what type of instrument is the most appropriate for a particular situation, how to write and test specific test items once the type of exam has been determined, what type of assessment method or scoring system should be used, who should be selected as readers and what type of training they should receive, what conditions are needed for administration of the exam, and how practical issues such as cost, administration sites, time concerns, and so on affect the testing process (1997, p.168).

Undoubtedly, validity and reliability constitute “the two cornerstones of the usefulness of a test” (Miguélez, 1997, p. 182). If a reliable measure is one that measures a construct consistently across time, individuals, and situations, then reliability can be said to relate directly to the consistency of test results. Validity refers to the degree to which a test or other examination instrument is truly measuring what it is supposed to measure. In designing tests for 18 language pairs, how does the National Center for State Courts assure that it is effectively incorporating the following five types of validity which are commonly recognized as essential to the construction of certification exams?

To begin, *content validity* has to do with how well items on a test align with what the test is supposed to measure. Consortium exams, as previously discussed, are based on real court documents and transcripts, are conscientiously normed according to the *Test Construction Manual*¹², and contain all three modes of interpreting that are used by practicing professional court interpreters. Furthermore, Consortium oral exams are considered to be task authentic precisely because they assess a candidate’s ability to apply standard-driven knowledge and skills to the real-world challenges of replicating the interpreting tasks which would be required of them in genuine professional settings.

Content validity, according to Giambruno (Miguélez), “overlaps somewhat with construct validity since construct validity has to do with how well an exam reflects the theoretical underpinnings that form its foundation” (1997, pp. 184-185). *Construct validity* refers to an assessment of how one’s (operationalized) ideas or theories have been translated into actual measures. In other words, does Consortium oral exam content actually measure what it proposes to measure? If one thinks in terms of aptitude for certain skills as generalized constructs, and the test in question tends to be consistent with other recognized measures of that construct, then the test has construct validity as a measure of aptitude. Test authenticity and task authenticity, however, require that a summative (versus formative) test functioning as a gateway to the profession examine and assess exactly those skills that the examinee is expected to perform successfully in the profession for which he or she is being tested. It is no surprise, then, that both certification systems used for spoken-language interpreters in the United States test the three modes of interpreting, with oral exams consisting of exercises in simultaneous interpreting, consecutive interpreting, and sight translation. Indeed, task authenticity implies quality, as “With a profession as practical and hands-on as interpreting, both teaching and examining should have a very strong focus on the work interpreters will

Competency-Based Education and Assessment

be doing...” (Heimerl-Moggan, 2016, p. 111). Consortium oral exams boast content validity because of the appropriateness of their use of actual court transcripts or written documents presented in court, and also because the test texts derived from these transcripts are reviewed by experienced professionals (including judges, lawyers, and interpreters) who have consistently found them to be representative of the language typically encountered in court.

Concurrent validity demonstrates that a test is valid by comparing it with other tests that are similar in purpose, format, scoring, methods, etc. A detailed report was commissioned by the Judicial Council of California’s Administrative Office of the Courts and carried out by ALTA Language Services, Inc.. Published in 2010 and entitled “California’s Assessment of the Consortium for Language Access in the Courts’ Exams”, the purpose of the California assessment was to determine the level of functional equivalency between California’s court interpreter certification examinations and the Consortium’s exams, and their detailed and well-documented analysis serves the purposes at hand of deepening our understanding of the way Consortium exams are constructed, rated, and administered, with the main principles of testing theory being addressed. The California assessment team was thoroughly satisfied that “[m]odeling the Consortium oral exams after the Federal oral exams, which have legally been shown to be valid (Arjona, 1985), provided the Consortium with a strong foundation for oral test development...” and that the Consortium’s

development and use of the TCM (Test Construction Manual) and an objective scoring method has further facilitated the development of a consistent and transparent method of achieving test validity during each stage of test development for every language for which the Consortium develops an interpreter certification test (ALTA Language Services, Inc., 2010, p. 4).

Indeed, the California assessment was most laudatory of the Consortium’s *Test Construction Manual*, declaring that:

[t]he TCM is a key element in establishing test equivalency across versions and languages by ensuring that each oral exam has the same distribution of scoring units, plus or minus 10 percent. The transparency and consistency as evidenced by the use of the TCM during every phase of Consortium test development contributes to a documented method of establishing test validity and equality (ALTA Language Services, Inc., 2010, p. 19).

In addition to the transparency, consistency, validity and equality referenced by the preparers of the California assessment, Consortium examinees are tested with instruments which have met the criteria to be considered authentic tests. In other words, candidates are asked to apply their inherent skills and knowledge, and their scores provide direct evidence of whether they can or cannot perform the skills that are being assessed: the three modes of interpreting.

Face validity has to do with stakeholders’ perceptions of the validity of an exam, or even with the perceptions held by society at large. Without face validity, a test may lose credibility and the support and participation of key sectors of society (van Deemter, Maxwell-Hyslop, & Townsley, 2014, p. 28¹³) with the potential result of a lack of trust in the system at large (Giambruno, 2014, pp. 188-189). In order to be considered valid and legitimate, the certifying entity or authority and, by extension, the exams they administer, must satisfy all stakeholders, be they governmental, educational, professional organizations, or private clientele. These stakeholders “rightfully expect certification (or credentialing) procedures to

be valid and reliable, and consider these authorities accountable to them” (Vermeiren, Van Gucht, & De Bontridder, 2009, p. 299). While neither the California assessment report nor the Consortium policy manuals address the issue of face validity, research has not revealed any controversy surrounding the way Consortium exams or policies are viewed by any group and, in fact, on the whole Consortium exams are held to be face valid precisely because of their concurrent validity (that is to say, their association with the federal certification exam), their content validity, and the authenticity of the tasks which examinees are asked to perform.

Finally, *predictive validity* is the extent to which a score on a scale or test predicts scores on some other criterion measure. However, the only studies in existence which gauge any aspect of predictive validity of Consortium oral exams as a whole or in part are a series of New Jersey predictor mode studies (Wallace, 2013) and the analysis of a large body of Consortium exam score data set which formed part of an unpublished 2012 doctoral dissertation entitled “Predictors of Successful Performance on U.S. Consortium Court Interpreter Certification Exams” (Wallace, 2012)¹⁴.

Room for Improvement: Limitations to Performance-Based Credentialing

As has been established, Consortium oral exams promise a high degree of validity and reliability from a psychometric perspective. Nonetheless, these exams feature limitations in terms of exactly what they assess. Such limitations are comprised of the dispositional traits and soft skills, as well as the substantive and technical knowledge, skills and abilities which are not assessed on oral certification exams, as shall be discussed below.

Knowledge, Skills and Abilities Not Assessed on Consortium Oral Certification Exams

Although exam constructs are sometimes ill-defined, for the purposes of an analysis of the Consortium exams, which are all constructed in accordance with the *Test Construction Manual*, the KSAs identified by the National Center for State Courts as essential to court interpreters consist of linguistic, speaking, listening comprehension, reading comprehension, and interpreting skills¹⁵ and do not measure

other aspects of the knowledge, skills, and abilities one must have to perform the duties of a court interpreter, such as engaging in appropriate forms of situational control, dressing and conducting oneself in a manner consistent with the dignity of the court, and delivering services via telephone or as a member of a team (National Center for State Courts, 2014, p. 1).

Scoring units, which are the elements assessed on an exam, cover the various types of language phenomena that interpreters must be able to correctly interpret such as numbers, names, legal terminology, general vocabulary, words used as modifiers and for emphasis, indicators of register or style, and elements of grammar and correct verb usage. In fact, Consortium oral certification exams rely *only* on objective scoring units. In an effort to support objective assessment practices, the Consortium exams do not

allow for an evaluation of subjective components, such as the testing of a candidate’s pronunciation or accent in the foreign language. In fact, the Consortium considers scoring elements such as foreign language accent difficult to test in an objective manner and therefore specifically avoids testing KSAs

Competency-Based Education and Assessment

(knowledge, skills and abilities) that cannot be measured through objective scoring (ALTA Language Services, Inc., 2010, p. 4)¹⁶.

Prosodic elements and delivery criteria such as fluency, intonation, a pleasant voice, native accent and diction are examples of constructs that are considered “largely intuitive” and difficult to evaluate due to their subjective nature (Iglesias Fernández, 2013). Iglesias Fernández identifies such constructs as “elusive” and “notoriously hard to define” (p. 52) even though they are often part of the evaluation criteria of certification exams. Such is not the case of the Consortium exams, which focus almost exclusively on linguistic aspects of examinees’ renditions, intentionally excluding pragmatic, para-linguistic and extra-linguistic factors. Indeed, while pronunciation and delivery seem to be inherent to language production in general, certain KSAs from the “speaking skills” and “listening skills” categories (as depicted in Table 1) are excluded from Consortium exams. Even if a candidate grossly mispronounces a word that is considered to be an acceptable rendition according to the scoring dictionary, if both raters understand the examinee and agree that what he or she has said is on the list of acceptable renditions, the examinee receives credit for that scoring unit (ALTA Language Services, Inc., 2010, p. 47).

A conscientious interpreter education program should involve attention to and formative assessment in these and other delivery-related competencies including fluency and voice quality (Iglesias Fernández, 2013), in addition to addressing the substantive knowledge limitations that are also characteristic of Consortium exams, as examinees are given little if any opportunity to demonstrate competence in certain domains such as ethics and ethical decision-making. This is in spite of the fact that

[t]here is widespread agreement that candidates should demonstrate knowledge of ethical standards in order to be certified, but the multiple-choice, machine-scorable written exams that have been chosen by many testing entities as the most cost-effective way of testing that knowledge are not universally recognized as appropriate instruments for measuring that knowledge (Mikkelson, 2013, p. 69).

Additionally, examinees are currently untested in the target as well as the “home” legal systems, knowledge which includes an understanding of elements such as the formal and informal terminology used in both languages. A competency-based educational program, in addition to providing exposure to interpreting techniques, codes of conduct, best practices, and professional organizations (Corsellis & Vanden Bosch, 2014, p. 69), could also lend attention to the soft skills and dispositional traits that are unassessed on performance-based exams but which contribute to interpreting quality. The next section surveys such skills and traits and makes the argument that they should no longer be overlooked in interpreter training and assessment.

Table 1. KSAs not assessed on Consortium exams

Domains	Speaking Skills (for All Working Languages)	Listening Skills (for All Working Languages)
Subdomains	2A: Proper pronunciation, diction, and intonation 2B: Neutralized accent 2C: ability to project and/or speak softly	3B: Listen to and comprehend regional accents and dialectal differences 3C: Ability to ignore auditory distractions and focus on source speaker

Source: ALTA Language Services, Inc., 2010, p. 47.

Soft Skills and Dispositional Traits

One might presume that a certification exam for court interpreters should consist solely of a performance examination in which the examinee is asked to show his or her competence in the three modes of interpreting and, indeed, few would question the necessity of these task performance elements. Nonetheless, “a valid examination may well involve more than a performance component. All aspects of competence must be taken into account” (Koby & Melby, 2013, p. 178). Soft skills, for example, are the intangible abilities of effective professionals, the personal attributes and dispositional traits that allow some people to work harmoniously and even thrive in a given setting. They may include aptitude in areas such as creativity, analytical thinking, multitasking, time management, or social graces and tactful communication skills, among others.

Iglesias Fernández well observes that “As the discipline advances from linguistic and pragmatic-discursive to socio-cultural perspectives, and interpreting grows exponentially in other settings, additional quality constructs, such as communicative, interpersonal and cross-cultural competences should be incorporated to broaden the notion of interpreter competence” (Iglesias Fernández, 2011, p. 13). Studies by Timarová and Salaets show preliminary evidence that in the field of interpreting, soft skills such as motivation and cognitive flexibility, often recognized as important but not systematically tested or researched, may be significant contributing factors in interpreting aptitude (2011, p. 49). In a similar vein, Koby and Melby affirm that in addition to the knowledge, skills and abilities which are technical or linguistic in nature and which form the constructs of the Consortium oral exams, “it is clear that inherent personal attributes and abilities of a competent professional are the basis for integrating knowledge and skills in a particular context in order to accomplish needed tasks” (2013, p. 179). In the same article, Koby and Melby produce an inventory of attitudinal traits that practicing professionals identified on a job task analysis survey done on behalf of the American Translators Association. In addition to the 36 KSAs (knowledge, skills, and abilities), the job task analysis included six personal attributes, ranked by relative usefulness or importance:

1. Thorough, meticulous, attentive-to-detail
2. Persevering, reliable, trustworthy, sense-of-integrity
3. Desire-to-excel, curious, willing-to-learn, intellectually honest
4. Open-minded, objective, accepting-of-criticism
5. Diplomatic, professional manner, respectful, culturally sensitive
6. Team-player, collegial, collaborative (Koby & Melby, 2013, p. 185).

In point of fact, however, the testing and accreditation models used by most court interpreters who access the profession, the state-level Consortium exams, do not account for attitudinal and dispositional traits at all. Soft skills and attitudinal characteristics such as those mentioned above could potentially be integrated into combined competency testing modalities, a proposal that will be developed below.

THE CASE FOR COMPETENCY-BASED EDUCATION AND ASSESSMENT FOR US COURT INTERPRETERS: A POSSIBLE NEW PARADIGM FOR CREDENTIALING

One of the primary concerns of this volume is with how a greater degree of consistency, certainty and reliability can be achieved in the assessment of interpreter performance. To that end, the chapter argues that competency-based education and assessment carried out in collaboration with input from professional stakeholders leads, necessarily, to the task validity and reliability that is so highly prized in gateway exams in the translational professions. The following discussion features a brief inventory of areas of competency that various experts and researchers have identified as desirable or necessary in the education of aspiring court and legal interpreters. In acknowledgement of the vital role that assessment plays in both professional and academic settings, competency-based education focuses on the formative function of assessment as well as the summative function. Given the case made regarding the explicit strengths and weaknesses of performance-based credentialing for court interpreters in the US context, and having identified potential pitfalls of one-off, high-stakes performance-based accreditation exams, the value of complementing them with models of competency-based education and assessment is explored.

Competency-Based Education for Court Interpreters

Competency-based models of education and assessment are relatively new concepts in American higher education. Competency-based Education, or CBE, is characterized primarily by awarding credit based on student learning, not time spent in class (McClarty & Gaertner, 2015, p. i), allowing students to progress through the curriculum at their own pace as soon as they have demonstrated mastery of a competency or competency cluster. This model may shorten the time to completion of a program of study as well as providing “a promising framework for more consistent outcomes” as opposed to academic programs which deliver uneven results (Jolley, 2017, p. 15). For court interpreters-in-training, they offer the further potential advantage over traditional training programs of complementing performance-only certification exams with proof of mastery of specific competencies, thus sending a clearer signal of what examinees know and are able to do. Iglesias Fernández acknowledges widespread consensus on the fact that

evaluation in interpreting training should follow a migration, starting with assessment of the process at the early and formative stages; moving towards a quality assessment of the product in mid course, as a diagnostic instrument allowing us to inspect the strategic, pragmatic, cognitive and interactional components of interpreter behaviour; and finally focusing on the product, seen as a service to users, in line with market norms (2011, p. 12)¹⁷.

The advantages of developing a well-designed competency-based educational model are succinctly summarized by Jolley, who suggests that such a model has the potential to improve upon traditional university-level programs by:

- Involving industry professionals in the development of competencies to ensure that what programs emphasize are up to date.
- Using backward design principles to build programs focused on relevant learning goals.

- Using rigorous assessments to reliably determine what candidates are able to do in real-world contexts
- Focusing instruction and learning activities on topics that prepare students for assessments.
- Recognizing skills incoming candidates have already mastered to reduce the time it takes to complete the program.
- Allowing students to move through at their own pace, potentially reducing costs. (2017, p. 17).

In keeping with the idea of progressive models of comprehensive training which feature ongoing feedback, in interpreter education modules, decisions must be made in early stages of test development regarding whether discrete competency testing is preferred to combined competency testing, which combines the skills that can be tested in the course of one activity¹⁸. The discrete competencies tested on the consortium exams are the KSAs – interpreting skills that can be measured by scoring units.

This chapter urges the exploration of incorporating a program of competency-based education as an educational pre-requisite to performance-based certification exams. While acknowledging that an educational program of this type may not address all competencies, it could address a broader range of them, including domain-specific knowledge, specialized terminology, and communication skills required for accurate language mediation.

Competency-Based Assessment in Interpreter Education

Competency-based assessment in interpreter education can comfortably accommodate aspects of an interpreter's performance which cannot be easily assessed in a certification scheme that relies fully on scoring units, including "inter-personal skills of explanation and delineation of the interpreter's role, bi-cultural proficiency in pragmatics and proxemics, adaptation to different speakers' idiomatic and discourse features, (and) interaction management..." (Hlavac & Orlando, 2015, p. 310), in addition to the other soft skills previously discussed. "Modern" assessment strategies, according to Vigier Moreno and Valero Garcés, "are able to move beyond error assessment -- beyond linguistic, structural, and lexical domains – in order to account for elements related to pragmatics, discourse, socio-cultural context, (and) power relations" (2017, p. 14)

Erik Hertog underscores the benefits of ongoing, continuous assessment by suggesting the development of a rubric containing participation and involvement in class (or in the online platform) including preparation of assignments and understanding of issues discussed taking into account both knowledge and integrated awareness components, the quality of oral presentations, the quality of written assignments, and the quality of practical assignments such as role-plays, observations, or a supervised practicum (Hertog, 2014, p. 80)¹⁹. Indeed, competency-based education favors assessment tools such as group projects, research papers, presentations, portfolios, and role-plays in simulated environments. To its potential peril, nevertheless, competency-based assessment is often holistic, and thus open to worrisome degrees of subjectivity (see Vigier Moreno & Valero Garcés, 2017), potentially leading to questionable reliability and inter-rater inconsistency if raters are not properly trained in their task (Mikkelson, 2013, p. 69). The obvious advantage of holistic assessment, however, is its role in formative assessment for training future interpreters for authentic, on-the-job tasks (Vigier Moreno & Valero Garcés, 2017; Garant, 2009; Mobaraki & Aminzadeh, 2012). A high degree of subjectivity can lend itself to a lack of trust by stakeholders, threatening the face validity of gateway exams due to their potential for impartiality. Buy-in from potential students accustomed to purely performance-based exams in the United States is a conceivable pitfall that would need to be addressed.

PROPOSAL OF A HYBRID MODEL FOR US COURT INTERPRETER CERTIFICATION

A hybrid credentialing system combining the reliability, validity and objectivity of current Consortium oral exams in tandem with training and formative assessment in technical, prosodic, communicative, intercultural and interpersonal competencies, in addition to certain areas of substantive knowledge, would allow for the testing of conceptual and theoretical knowledge as well as interpreting skills in court and legal settings.

Specifically, the hybrid model would consist of a series of competency clusters which could be offered in person or online in a language-neutral environment, allowing examinees to receive feedback and training in (as well as recognition for) the main competencies not contemplated by Consortium oral exams. These modules would in no way be meant to replace the final oral exams but rather to complement them, with the idea being that a more complete and nuanced credentialing scheme leads to a higher quality testing experience. This “complete package” of competencies would offer examinees the option to complete any or all competency clusters in combination with a Consortium oral exam in their language pair, and it could, furthermore, offer students of court interpreting the possibility to work at their own pace and allow them to test out of “modules” in which they already have competency, much like the Finnish vocational model (see Wallace, 2017).

CBE models can take a variety of forms, but most programs include two common elements: (1) a competency framework and (2) competency assessments. The competency framework describes the “skills, abilities, and knowledge needed to perform a specific task” (McClarty & Gaertner, 2015, p. 2), and since those are almost completely contemplated in the Consortium oral exam, some parts of the competency-based training and assessment could and should focus specifically on the KSAs identified as essential for court interpreters (ALTA Language Services, Inc., 2010, p. 43), as well as the speaking and listening skills which are *not*, as detailed previously in Table 1.

Table 2 features some of the soft skills including intellectual and attitudinal competencies which would complement the skills tested on Consortium exams. Like the other competency clusters proposed in this hybrid model, students could be assessed on the basis of individual or group projects, research papers, presentations, portfolios, role plays, or other means. Substantive knowledge could also be addressed through a module dedicated to ethics and ethical decision-making in court and legal settings (Mikkelsen, 2013; Hertog, 2014) as well as legal concepts and knowledge of legal systems (Corsellis & Vanden Bosch, 2014). Similarly, a broad understanding of the field and its professional culture, as well as the practices and procedures of allied professions would be valuable.

Table 2. *Dispositional and intellectual competencies not assessed on Consortium exams*

Domains	Intellectual Competencies and Academic Skills	Dispositional and Attitudinal Traits
Subdomains	<ul style="list-style-type: none"> • Cognitive flexibility • Intellectual curiosity; open-mindedness • Accepting of criticism • Motivation; willingness to learn • Objectivity • How to research and prepare for assignments • Ability to engage in reflective practice • Knowledge of theories that underpin practice 	<ul style="list-style-type: none"> • Perseverance; resilience • Integrity; intellectual honesty • Diplomacy; professionalism; respectfulness • Collaborative nature; team player

Sources: Timarová & Salaets, 2011; Koby & Melby, 2013; Corsellis & Vanden Bosch.

Perhaps best suited to the competency-based education model, a cluster dedicated to communicative, interpersonal and cross-cultural competencies, such as that suggested in Table 3, would significantly add to the level of professionalism and awareness evinced by novice interpreters upon entering the field:

Certainly competency-based credentials, in order to enjoy perceived value and face validity, would depend fully on the reliability and validity of their assessment tools. Standard setting processes and the identification of cut scores for each assessment type would need to be determined based on data indicating the level of mastery necessary in order to demonstrate adequate or minimal competency. There would have to be an explicit link between the competencies identified as necessary and the skills measured by the assessment instruments. Content would have to be evidence-based and instruments would have to be task-authentic and overall must be an authentic reflection of all the competencies that court interpreters entering the profession would be expected to have in order to deliver quality interpreting services. Certainly, psychometric norming of the assessment tools themselves would be vital, as would a conscientious development of protocols for rating, monitoring, item testing, security, and rater selection and training.

Norway has a system similar to the one envisioned in this chapter, with a national register (established in 2005) that includes categories that represent entryways to the register rather than levels of qualification. The five categories of interpreters can be summarized as (1) interpreters with university-level training AND the Norwegian Interpreter Certification Exam; (2) interpreters with the Norwegian Interpreter Certification Exam; (3) interpreters with completed university-level training; (4) people with a BA or a state authorization test in written translation and a three-day course on professional ethics; (5) people who have passed a limited bilingual lexical knowledge test and completed a three-day course on professional ethics (Giambruno, 2014, pp. 178-180). The idea behind this “stepping stone” system is to inspire practitioners to advance in the system by continuing to enhance and document their qualifications. In the context of the United States, a tiered register could not realistically include required university-level training in interpreting at this time due to the paucity of programs in higher education, and traditionally the idea of requiring a minimum threshold of formal education for examinees has not been embraced. However, they might include categories such as those in Table 4:

Table 3. Communicative, interpersonal, and cross-cultural competencies not assessed on Consortium exams

Domains	Communicative and Interpersonal Competencies	Cross-Cultural Competencies
Subdomains	<ul style="list-style-type: none"> • Situational control; interaction management • Ability to navigate formal and informal registers • Ability to delineate the interpreter’s role • Ability to adapt to discourse features of different speakers 	<ul style="list-style-type: none"> • Bi-cultural proficiency in pragmatics and proxemics • Cultural sensitivity and awareness

Sources: Koby & Melby, 2013; Iglesias Fernández, 2011; National Center for State Courts, 2014; Corsellis & Vanden Bosch, 2014; Hlavac & Orlando, 2015.

Table 4. Proposed hybrid model for court interpreter certification

	Tier 1	Tier 2	Tier 3
Passed Consortium oral exam	✓		
Passed Consortium written exam (if applicable; varies by state)	✓	✓	
State-mandated orientation course (if applicable; varies by state)	✓	✓	✓
Completed and passed competency clusters (language-neutral)	✓	✓	✓

Competency-Based Education and Assessment

Indeed, in light of the reality that a minimum threshold of education for examinees is not currently under consideration, the target population of potential examinees would not vary from what it is now, remaining essentially open to everyone and thus making access to training for interpreters of languages of lesser diffusion a priority. The proposed modules discussed in this section would, ideally, constitute pre-requisites to the current oral exams on offer, and they could be offered by educational institutions in direct collaboration with the National Center for State Courts.

DISCUSSION

This chapter delineates how current certification schemes may not necessarily test all of the knowledge, skills and abilities that court interpreters need to possess in order to practice effectively and accurately, especially in light of the dearth of educational opportunities available to them at the current time. The fact that neither the FCICE nor the Consortium exams have any educational requirement at all (Paradigm Testing, 2017, p. 2; National Center for State Courts, 2006, p. 1)²⁰ highlights the underdeveloped nature of the connection between training and testing²¹, thus making the proposal for educational clusters relevant to education as well as certification.

By combining the best elements of at-your-own-pace competency-based instruction and assessment modules in an online, language-neutral format, a wider net can be cast for prospective court interpreters in many more languages. The subsequent creation of a multi-tiered register may contribute to professionalization efforts. A model like the Norwegian one, or the one proposed herein, increases meaningful training and credentialing options for speakers of languages of lesser diffusion for whom oral certification exams are not available. It would allow practitioners to test out of or study for and pass the assessments for competency-clusters that they can demonstrate they have mastered, and it would continue to place primacy on the psychometrically sound Consortium oral exams already in existence in 18 languages.

Competency-based assessment may be considered innovative in the context of court interpreter certification in the United States and would require the transfer of assessment criteria usually used in traditional educational contexts into the realm of professional training, with new kinds of feedback and a shift from one high-stakes assessment experience to a series of compartmentalized competency clusters, in contrast to a historical reliance solely on performance-based accreditation. As advocated by Hlavac and Orlando, it would contribute towards the creation of a culture of on-going self-reflection, self-development and self-improvement (2015, p. 321). By making the competency-based clusters available in a language-neutral format, quality training could be extended to emerging interpreters of languages of lesser diffusion who often serve particularly vulnerable populations.

Possible limitations to this proposal must be acknowledged, however. A training and testing scheme with so many parameters and so many stakeholders must strike a balance between test efficiency and the resources needed to administer it, especially as cost-effectiveness is a ubiquitous concern for language access programs worldwide. More research would need to occur in order to operationalize the proposed competency clusters detailed in Tables 2 and 3: in other words, important decisions regarding assessment types, measurement criteria, and rating mechanisms would need to be decided upon.

Inevitably, external validity would require data tracking and oversight which may not be practicable for the National Center for State Courts to oversee, although a collaboration between institutions of higher learning could conceivably build, staff, maintain, oversee and monitor the modules. NCSC buy-in would be an excellent way to legitimize the additional testing components.

In essence, performance-based credentialing, if done with psychometrically normed instruments that are reliable, valid, and task authentic, and which are reliably rated by trained and qualified raters, is a sound way to certify court interpreters, albeit with its limitations. Nevertheless, if combined with competency-based assessment and training, a hybrid model such as that proposed here may offer significant flexibility and could be implemented in the United States as a collaborative project developed between subject matter experts, academics, and hopefully the National Center for State Courts.

REFERENCES

- ALTA Language Services, Inc. (2010). *Beyond words: California assessment of the Consortium for Language Access in the Courts' exams*. Retrieved from <http://www.courts.ca.gov/documents/ALTAReport.pdf>
- Arjona, E. (1985). The court interpreters certification test design. In L. Elías-Olivares, E. A. Leone, R. Cisneros, & J. R. Gutiérrez (Eds.), *Spanish language use and public life in the USA* (pp. 181–200). Berlin: Mouton de Gruyter.
- Arjona-Tseng, E. (1994). A psychometric approach to the selection of translation and interpreting students in Taiwan. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation* (pp. 69–85). Amsterdam: John Benjamins. doi:10.1075/btl.3.08arj
- Corsellis, A., & Vanden Bosch, Y. (2014). Testing knowledge of the legal context. In C. Giambruno (Ed.), *Assessing interpreter quality through testing and certification: The QUALITAS project* (pp. 69–76). Sant Vicent del Raspeig, Spain: Publicaciones de la Universidad de Alicante.
- de Jongh, E. (2012). *From the classroom to the courtroom: A guide to interpreting in the U.S. justice system*. Amsterdam: John Benjamins. doi:10.1075/ata.xvii
- Feuerle, L. M. (2013). Testing interpreters: Developing, administering, and scoring court interpreter certification exams. *The International Journal of Translation and Interpreting Research*, 5(1), 297–323. doi:10.12807/ti.105201.2013.a04
- Garant, M. (2009). A case for holistic assessment. *A FinLA-e Soveltavan Kielittieteen Tutkimuksia*, 1, 5–17. Retrieved from <http://hdl.handle.net/10138/154545>
- Giambruno, C. (2014). The current state of affairs in the UE: Member state profiles. In C. Giambruno (Ed.), *Assessing interpreter quality through testing and certification: The QUALITAS project* (pp. 149–190). Sant Vicent del Raspeig, Spain: Publicaciones de la Universidad de Alicante.
- Giambruno, C. (2016). Quality assessment and political will. *Language and Law/Linguagem E Direito*, 3(2), 116–134.
- Giambruno (Miguélez), C. (1997). *Language mediation in the judicial system: The role of the court interpreter* (Unpublished doctoral dissertation). Universitat d'Alacant, Spain. Retrieved from file:///Users/dqf115/Downloads/Giambruno-Cynthia.pdf
- González-Hibner, M., Palma, J., & Korp-Edwards, B. (2017). *From the NAJIT board: The future of the NJITCE*. Retrieved from <https://najit.org/from-the-najit-board-the-future-of-the-njitce/>

Competency-Based Education and Assessment

Heimerl-Moggan, K. (2016). Fit-to-practise versus exam-ready legal public service interpreter training: Training for real life or for the exam? In T. M., G. Webb & M. Rabadán-Gómez (Eds.), *Challenges and opportunities in public service interpreting*. Academic Press. doi:10.1057/978-1-137-45000-5_5

Hertog, E. (2014). Testing knowledge of codes of ethics. In C. Giambruno (Ed.), *Assessing legal interpreter quality through testing and certification: The QUALITAS project* (pp. 77–92). Sant Vicent del Raspeig, Spain: Publicaciones de la Universidad de Alicante.

Hlavac, J. (2013). A cross-national overview of translator and interpreter certification procedures. *The International Journal of Translation and Interpreting Research*, 5(1), 32–65. doi:10.12807/ti.105201.2013.a02

Hlavac, J., & Orlando, M. (2015). ‘Capturing’ and ‘prescribing’ desirable attributes of community interpreting: Design and implementation of testing systems in four anglophone countries. In C. Zwischenberger & M. Behr (Eds.), *Interpreting quality: A look around and ahead* (pp. 297–324). Berlin: Frank & Timme.

Iglesias Fernández, E. (2011). Under examination. Do all interpreting examiners use the same criteria? *The Linguist*, 50-2. Retrieved from https://www.researchgate.net/publication/266317103_Under_Examination_Do_All_Interpreting_Examiners_Use_The_Same_Criteria

Iglesias Fernández, E. (2013). Unpacking delivery criteria in interpreting quality assessment. In D. Tsagari & R. van Deemter (Eds.), *Assessment issues in language translation and interpreting* (pp. 51–66). Frankfurt: Peter Lang.

Jolley, J. (2017, July/August). Can competency-based education enhance translator training? *The ATA Chronicle*, 46, 15-17. Retrieved from <http://www.atanet.org/chronicle-online/featured/can-competency-based-education-enhance-translator-training/#sthash.UyFrvLON.dpbs>

Koby, G., & Melby, A. K. (2013). Certification and job task analysis (JTA). Establishing validity of translator certification examinations. *The International Journal of Translation and Interpreting Research*, 5(1), 174–210. doi:10.12807/ti.105201.2013.a10

McClarty, K. L., & Gaertner, M. N. (2015). *Measuring mastery. Best practices for assessment in competency-based education*. Retrieved from <http://www.aei.org/publication/measuring-mastery-best-practices-for-assessment-in-competency-based-education/>

Miguélez, C. (1999, Spring). Current issues in court interpreting: Spain, a case study. *Proteus: Newsletter of the National Association of Judiciary Interpreters and Translators*, 8-2. Retrieved from http://www.najit.org/members_only/proteus/v8n2/miguelev8n2.htm

Mikkelsen, H. (2013). Universities and interpreter certification. *The International Journal of Translation and Interpreting Research*, 5(1), 66–78. doi:10.12807/ti.105201.2013.a03

Mobaraki, M., & Aminzadeh, S. (2012). A study on different translation evaluation strategies to introduce an eclectic method. *International Journal of English Linguistics*, 2(6), 63–70. doi:10.5539/ijel.v2n6p63

National Center for State Courts. (2006). *Court interpreting qualifications*. Retrieved from <http://www.ncsc.org/~media/Files/PDF/Education%20and%20Careers/State%20Interpreter%20Certification%20Qualifications.ashx>

- National Center for State Courts. (2016, November 9). *Court interpreter oral exams. NCSC oral examinations ready for administration (20 languages)*. Retrieved from http://www.ncsc.org/~media/Files/PDF/Services%20and%20Experts/Areas%20of%20expertise/Language%20Access/Resources%20for%20Program%20Managers/Oral_Exam_Ready_for_Administration%20November%202016.ashx
- National Center for State Courts. (2017, September 19). *NCSC oral examinations ready for administration*. Retrieved from http://www.ncsc.org/~media/files/pdf/services%20and%20experts/areas%20of%20expertise/language%20access/written%20and%20oral/oral_exam_ready_for_administration%20september%202017.ashx
- National Center for State Courts – Language Access Services Section. (2014). *Court interpreter oral examination: Test Construction Manual*. Retrieved from http://www.ncsc.org/~media/Files/PDF/Services%20and%20Experts/Areas%20of%20expertise/Language%20Access/Written%20and%20Oral/2014%20January_Test%20Construction%20Manual%201%202029%2014.ashx
- Niska, H. (2002). Community interpreter training. Past, present, future. In G. Garzone & M. Viezzi (Eds.), *Interpreting in the 21st century* (pp. 133–144). Amsterdam: John Benjamins. doi:10.1075/btl.43.14nis
- Ortega, J. M., Sandrelli, A., & Townsley, B. (2014). Interpreting test format. In C. Giambruno (Ed.), *Assessing legal interpreter quality through testing and certification: The QUALITAS project* (pp. 41–67). Sant Vicent del Raspeig: Publicaciones de la Universidad de Alicante.
- Ozolins, U. (2016). The long road to effective court interpreting. *Language and the Law / Linguagem E Direito*, 3(2), 98–115.
- Paradigm Testing, L. L. C. (2017, June 1). *Federal Court Interpreter Certification Examination for Spanish/English*. Retrieved from <https://paradigmtesting.com/fcice-hb/>
- Skaaden, H., & Wadensjö, C. (2014). Some considerations on the testing of interpreting skills. In C. Giambruno (Ed.), *Assessing legal interpreter quality through testing and certification: The QUALITAS project* (pp. 17–26). Sant Vicent del Raspeig, Spain: Publicaciones de la Universidad de Alicante.
- Stansfield, C. W., & Hewitt, W. E. (2005). Examining the predictive validity of a screening test for court interpreters. *Language Testing*, 22(4), 438–462. doi:10.1191/0265532205lt304oa
- Timarová, Š., & Salaets, H. (2011). Learning styles, motivation and cognitive flexibility in interpreter training. *Interpreting*, 13(1), 31–52. doi:10.1075/intp.13.1.03tim
- van Deemter, R., Maxwell-Hyslop, H., & Townsley, B. (2014). Principles of testing. In C. Giambruno (Ed.), *Assessing legal interpreter quality through testing and certification: The QUALITAS project* (pp. 27–39). Sant Vicent del Raspeig, Spain: Publicaciones de la Universidad de Alicante.
- Vermeiren, H., Van Gucht, J., & De Bontridder, L. (2009). Standards as critical success factors in assessment: Certifying social interpreters in Flanders, Belgium. In C. V. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting: A call for dialogue between research and practice, American Translators Association Scholarly Monograph Series* (Vol. 14, pp. 297–328). Amsterdam: John Benjamins. doi:10.1075/ata.xiv.14ver

Vigier Moreno, F. J., & Valero Garcés, C. (2017). Shedding light on the grey area of assessment in legal translation: A case study based on a holistic method. *FITISPos International Journal*, 4, 11-27. Retrieved from <http://dspace.uah.es/dspace/handle/10017/29800>

Wallace, M. (2012). *Predictors of successful performance on U.S. Consortium court interpreter certification exams* (Unpublished doctoral dissertation). Universitat d'Alacant, Alicante, Spain. Retrieved from https://rua.ua.es/dspace/bitstream/10045/28364/1/tesis_melissawallace.pdf

Wallace, M. (2013). Rethinking bifurcated testing models in the court interpreter certification process. In D. Tsagari & R. van Deemter (Eds.), *Assessment issues in language translation and interpreting* (pp. 67–83). Frankfurt am Main: Peter Lang.

Wallace, M. (2015). A further call to action: Training as a policy issue in court interpreting. *The Interpreter and Translator Trainer*, 9(2), 173–187. doi:10.1080/1750399X.2015.1051769

Wallace, M. (2017). Lessons from Finland: Diversity and flexibility in the authorization of court interpreters. In M. Kivilehto, A. Liimatainen, A. Nurmi, L. Salmi, A. Viljanmaa, & M. Wallace (Eds.), *Legal translation and court interpreting: Ethical values, quality, competence training (Forum für Fachsprachenforschung)*. Berlin: Frank & Timme.

KEY TERMS AND DEFINITIONS

Content Validity: Content validity is concerned with a test's ability to include or represent all of the content of a particular construct. In other words, it has to do with how well items on a test align with what the test is supposed to measure. There is no easy way to determine content validity aside from relying on expert opinion.

Face Validity: Face validity has to do with stakeholders' perception of the validity of an exam. Its importance lies in the fact that if practitioners and society at large do not perceive a tool or instrument as valid, it will not be accepted and its usefulness will be undermined.

Formative Assessment: Formative assessment is part of the instructional process. When incorporated into classroom practice, it provides the information needed to adjust teaching and learning while they are happening, informing both teachers and students about student understanding at a point when timely adjustments can be made, thus helping to ensure that students achieve targeted standards-based learning goals within a set time frame. Formative assessments tend to be low-stakes evaluations for students and instructors, and they are rarely (if ever) used in interpreter certification exams. While examinees receive a minimal amount of feedback regarding their performance on certification exams, such exams rely on summative types of assessment.

Inter-Rater Reliability: Inter-rater reliability refers to the degree of consensus between raters of an exam. In other words, an assessment tool with a high rate of inter-rater reliability is one in which all raters of the same assessment arrive at approximately the same results when evaluating a candidate so that passing does not depend upon who grades a specific exam. If various raters do not agree, either the scale they are using to evaluate their examinees is defective or the raters need to be retrained.

Predictive Validity: Predictive validity is the extent to which a score on a scale or test predicts scores on some criterion measure. For example, the predictive validity of an entry exam for student interpreters

(designed to measure their aptitude) is the correlation between entry test scores and, for example, final marks in their program of training or final grades on an exit exam. Tests which have predictive validity of the observed correlations are statistically significant.

Psychometrics: The field of study concerned with psychologically measuring human characteristics such as knowledge, skills, abilities, attitudes, and personality traits. Specialists in this field analyze and research the construction and validation of measurement instruments.

Reliability: In testing theory, a reliable measure is one that measures a construct consistently across time, individuals, and situations.

Scoring Units: Scoring units are the elements which are assessed on an exam. In the context of oral interpreting exams, they reflect various types of language phenomena that interpreters must be able to correctly interpret. On consortium court interpreting exams, different types of scoring units include numbers, names, legal terminology, general vocabulary, words used as modifiers and for emphasis, indicators of register or style, and elements of grammar and correct verb usage.

Summative Assessment: The goal of summative assessment is to measure the level of success or proficiency of an examinee by comparing it against some standard or benchmark. Oral certification exams are preeminently summative in nature given that they are of a pass/fail nature, and indicate to the examinee whether he or she meets a minimum standard of competency in the skill areas being tested. The outcome of a summative assessment can be used formatively; however, when students or faculty take the results and use them to guide their efforts and activities in subsequent courses. Oral certification examinees, for example, can interpret a failing grade on certain interpreting exercises to signify a lack of skills involved in the mode of interpreting on which they did not meet a minimal standard.

Task Authenticity: Task authenticity assesses a person's ability to apply standard-driven knowledge and skills to real-world challenges. In other words, the task replicates challenges faced in real professional settings. Oral interpreting certification exams which are considered task authentic ask examinees to perform the three modes of interpreting, as they are expected to use all three modes successfully in a professional context.

Test Authenticity: Test authenticity is closely related to task authenticity. An authentic test asks examinees to perform a task as opposed to selecting a response; test performance is not contrived or artificial, but can be considered similar to "real-life". An examinee cannot depend on recall or recognition to succeed on an authentic test, but must be able to apply inherent skills and knowledge. Authentic tests provide direct evidence that a candidate can or cannot perform the tasks that are being assessed.

Validity: Validity refers to the degree in which a test or other measuring device is truly measuring what it intends to measure. Five types of validity are generally recognized in relation to the construction of certification exams. See definitions of concurrent validity, construct validity, predictive validity, content validity, and face validity.

ENDNOTES

¹ Accreditation and certification are used interchangeably in this chapter.

² Consortium exams are not *federal* as they are not the ones designed for interpreting in federal court; nonetheless, they may be considered *national* in the sense that as of 2012, all US states and territories have access to the test bank, and there are reciprocity measures in place between states.

Competency-Based Education and Assessment

- ³ Read the Directive in its entirety at <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:280:0001:0007:en:PDF>.
- ⁴ Please note that the author cited here has published as Giambruno, Giambruno (Miguélez), and Miguélez. References to her research will reflect the last name(s) under which works were published.
- ⁵ This chapter offers a general overview of the two paths to certification currently available in the United States, and does not include detailed information about the development of the exams, their pre-requisites, their pass rates, or other details. Discussion has been limited to psychometric factors related to testing theory. Neither set of exams is accredited by the NCCA (National Consortium for Certifying Agencies) nor was a job task analysis conducted for either of them, to the author's knowledge.
- ⁶ The NAJIT (National Association of Judiciary Interpreters and Translators) certification exam, although considered prestigious and still accepted for state certification by Colorado, Connecticut, Delaware, Hawaii, Iowa, Massachusetts, New York, Pennsylvania, Rhode Island, Texas and Wisconsin, was only offered for the Spanish / English language. The last NAJIT exam was administered in 2012 (González-Hibner, Palma, & Korp-Edwards, 2017). NAJIT's board of directors felt that the cost-benefit ratio, and perhaps even the relevance of the NAJIT certification itself, had come into question once the state-level exams administered by the National Center for State Courts became available to all US states and territories in 2012.
- ⁷ Data regarding pass rates for the FCICE are not publicly available but are rumored to be somewhere between four and six percent.
- ⁸ The Council of Language Access Coordinators (CLAC) was chartered in 2012, essentially obliging all US states and territories to designate a point person to oversee language access services to the courts. Prior to this date, the Consortium of Language access in the Courts (also CLAC) existed as a multi-state partnership dedicated to sharing the expenses of developing and maintaining oral court interpreting exams under the auspices of the National Center for State Courts. See Wallace (2012) for more information about the formation of the Council of Language Access Coordinators.
- ⁹ Hildegard Vermeiren, Jan Van Gucht and Leentje De Bontridder contrast social interpreter testing in Flanders, Belgium, where candidates are tested strictly on competencies relevant to interpreter performance. Other countries such as Sweden or the United Kingdom assess interpreting performance as well as cultural and terminological knowledge (2009, p. 307).
- ¹⁰ Hlavac & Orlando (2015, pp. 302-304) provide a useful inventory of concepts of quality over time as applied to interpreting.
- ¹¹ Consortium exams in particular are summative, high-stakes exams which allow or impede professional practice in court. See also Giambruno, 2016 and Hlavac & Orlando, 2015 for a discussion of testing theory concepts as applied to court interpreter certification exams.
- ¹² Found at http://www.ncsc.org/~media/Files/PDF/Services%20and%20Experts/Areas%20of%20expertise/Language%20Access/Written%20and%20Oral/2014%20January_Test%20Construction%20Manual%201%2029%2014.ashx.
- ¹³ See van Deemter et al., 2014 for a discussion of norm-referenced versus criterion-referenced certification exams.

- ¹⁴ Stansfield & Hewitt (2005) examined the predictive validity of the *written* part of the Consortium certification exams.
- ¹⁵ The domains and subdomains of the KSAs which the Consortium considers essential for court interpreters are detailed on page 43 of ALTA Language Services' report entitled *Beyond words: California assessment of the Consortium for Language Access in the Courts' exams*.
- ¹⁶ See ALTA Language Services, Inc., 2010, p. 47 for a list of the KSAs for court interpreters that are directly measured on Consortium oral exams.
- ¹⁷ For an inventory of interpreter training and accreditation standards as of 2013 in 21 countries, see Hlavac, 2013).
- ¹⁸ See examples of approaches to discrete versus combined competency testing in countries such as Belgium and the UK (Ortega et al., 2014, pp. 64-65).
- ¹⁹ See Hertog 2014 for specific ideas about how knowledge of ethics could be incorporated into role-plays as part of a competency-based assessment.
- ²⁰ The National Center for State Courts asks exam candidates to self-evaluate their suitability to take the written and oral certification exams, advising that professional court interpreters are individuals who “display wide general knowledge characteristic of what a minimum of two years of a general education at a college or university would provide” (2006, p. 1).
- ²¹ To the author's knowledge, no information is publicly available that would shed light on the professional training or formal education held by test candidates. The FCICE does request demographic information such as this when a candidate enrolls to begin the federal testing process but it has not made the information public.

Chapter 6

Dischronies and the Absence of Image in the Evaluation of Dubbing

José Tomás Conde Ruano
University of the Basque Country, Spain

ABSTRACT

An approximation of quality in audiovisual translation is presented, based upon an experiment carried out with students of different level of expertise, in order to check whether considering the image affects the evaluation of quality in dubbing and which dischrony has a greater impact on students' ratings. When they watch the scenes on video, evaluators (i.e. the students) bestow them higher ratings as opposed to when they have only the text. The lack of synchronization, which is penalized the most, is lip synchronization, although the differences do not seem to be significant. The average ratings given by the two groups of students are similar, but there is a stronger coincidence among advanced students. The presentation on video beclouds potential linguistic errors made during the translation phase. Errors related to synchronization are not as relevant to the receivers' judgement as expected.

INTRODUCTION

'Quality' and 'evaluation' are recurring concepts in the literature on translation and interpreting. While earlier studies addressed such concepts on a fairly intuitive manner and by means of fundamentally prescriptive approaches, the 21st century has witnessed a boom of empirical research resulting in more up-to-date and comprehensive knowledge on the feasibility of assessment methods and tools (Waddington, 2001) and what is meant by a 'good' or 'bad' translation (Conde, 2009). In the field of audiovisual translation (AVT), however, only a few studies have dealt with quality and evaluation of audiovisual products, and sometimes the scope of this research is very limited.

The study of expertise in translation is also to the fore in the literature on translation studies (see, for instance, the work carried out by the PETRA Research Group). Translation expertise is usually investigated on the grounds of translation process research and consists of a comparison of the way experts

DOI: 10.4018/978-1-5225-5225-3.ch006

and non-experts behave, not only when translating, but also towards other tasks and skills that are at the core of the translation activity, such as reading, problem-solving or assessing.

This chapter is a first approximation to the study of the evaluation of AVT with a scientific methodology and taking into account the potential differences in two groups of students with different levels of expertise (i.e. second- and fourth-year undergraduate students). It is a preliminary study aiming at discovering how important is for the assessment of AVT the format in which the translation is presented and which type of dischronies has a greater impact on the evaluation of the dubbing quality.

BACKGROUND

AVT products, by matching text and image, do not evoke but represent a concrete reality that restricts translation (Díaz-Cintas, 2003, p. 33). The image takes part in the creation of meaning, being crucial for AVT professionals to have both the scripts and the video at their disposal (Chaume, 2004, pp. 268-269; Segovia, 2005, p. 85; Sokoli, 2005, p. 185). Some research (Chaume, 2004, p. 89; Díaz-Cintas, 2003, p. 125) shows, however, that this is not always the case.

De los Reyes (2015, pp. 28-29) gathers the criticisms about studies that do not focus on aspects such as code interaction and synchrony. Another pertinent question to be asked is whether having the video or not also influences the response to the audiovisual texts. Image and text interaction is common to all the modalities of AVT, although the present chapter focuses on the most common modality in Spain (Chaume, 2004, p. 33): dubbing.

Dubbing

According to Chaume (2012, p. 1) dubbing consists of ‘replacing the original track of a film’s (...) source language dialogues with another track on which translated dialogues have been recorded in the target language’. Much has been written about the differences between dubbing and other modalities of AVT; each modality has advantages and disadvantages as well as particular challenges (Díaz-Cintas, 2003, p. 294).

From a theoretical point of view, Hurtado (2001, p. 76) considers dubbing a type of a simple subordinated translation, since it maintains the medium of the original text. From a practical point of view, Agost (1999, p. 97) claims that the feature that singles out the dubbing modality is the synchrony.

‘Adjusting’ or ‘synchronising’ consists of ‘matching the target language translation and the articulatory and body movements of the screen actors and actresses, and ensuring that the utterances and pauses in the translation match those of the source texts’ (Chaume, 2012, p. 68).¹ Given this element is paramount for dubbing, it is no wonder that synchronising has been addressed so profusely. Most of the studies that deal with this aspect in dubbing suggest that synchronies can be of three types (Agost, 1999, p. 65; Chaume, 2012, pp. 68-69; Arumí et al., 2013, p. 47):

- Lip synchrony (henceforth, L): between the lip movement and the words that are articulated;
- Kinetic synchrony (henceforth, K): between the body movements and what is heard; and
- Isochrony (henceforth, I): between the duration of the utterances and pauses.

Dischronies and the Absence of Image in the Evaluation of Dubbing

Synchronies are more than a simple characteristic of dubbing. Mayoral (2005, p. 6) states that adjusting is the central issue for the study of AVT, for it has a huge effect on the quality of the final product and because it refers to most of its theoretical problems. For example, Weidmann (2014, p. 65) links the nature of scenes or shots with the level of acceptance of the three types of synchronies. Furthermore, Chaume (2004, p. 256) lists several factors that have an impact on the reception of synchronies, such as similarity of the original language and the target language, cultural differences, conventions, types of clients or audiences or the screen – TV or cinema – on which the product might be presented (Chaume, 2012, pp. 76-78).

Thus, it seems that synchrony is closely related to the quality assessment of audiovisual translated products.

Quality

Some studies have dealt with the concepts of quality and evaluation in specific modalities of translation (either dubbing or subtitling), with an increasing interest and concern about quality in AVT over the last twenty years.²

Despite focusing on a different modality (i.e. subtitling), according to Bittner (2011, p. 86) the quality of a translation is crucial for those films aiming to be considered a work of art. Díaz-Cintas (2003, p. 71) agrees with Fuentes Luque (2005, p. 139) in that the success of an audiovisual work corresponds to the quality of the translation of the text and, in the case of dubbing, more so to the quality of the synchrony.

Quality in AVT might depend on factors as diverse as economic circumstances or time constrains (Agost, 1999, pp. 61-62; Abdallah, 2012, p. 33; Orero, 2005, p. 213); the tools available (Orero, 2005, p. 214); the revision process (Díaz-Cintas, 2003, p. 82) and other phases (Díaz-Cintas, 2003, p. 86); the existence of internal policies (Alfaro, 2012, pp. 464-465); the rationalisation of work and the lack of experience of the professionals (Díaz-Cintas, 2003, p. 72).

Chaume (2005, p. 6 and 2007, pp. 74-75) suggested a series of parameters to measure quality in dubbing:

- Respect for the mouth and body movements,
- Image-text coherence,
- Fidelity to the original,
- Dramatization and
- Technical conventions.

Some authors have proposed some tools for quality control, such as corpora of scripts (De Yzaguirre, García, Matamala Ripoll, & Ribas, 2005, p. 109); classification systems agreed by different groups, such as translator unions and associations (Abdallah, 2012, p. 48); style guides (Alfaro, 2012, p. 470) and more or less objective scales, such as those introduced by De los Reyes (2015, p. 187) or Jekat, Prontera, and Bale (2015, pp. 448-450).

As Díaz-Cintas (2003, p. 204) states about subtitling, despite the quality of audiovisual translation seems to rely on the overall perception of the spectator, 'no empirical evidence has shown what a good dubbing is' (Chaume, 2012, p. 15). Nonetheless, synchrony seems to play an important role when spectators evaluate AVT. Indeed, Weidmann (2014, p. 68) refers to synchronism as one of the key criteria to assess the quality of dubbing. Taking synchrony as an example, Chaume (2005, p. 6) considers that anything

that goes against customary conventional norms is perceived as a mistake by the receiver; according to Agost (1999, p. 140), spectators consider the interpretation of the actors and visual synchrony. As a matter of fact, according to Mayoral (2001, p. 36) the quality in AVT is linked to the visual synchrony.

Far beyond the contributions above, the research on the evaluation of translations exposes the significance not only of the synchrony itself, but also the receiver (Di Giovanni, 2016, p. 77; Abdallah, 2012, p. 36; Alfaro, 2012, p. 470; Bucaria & Chiaro, 2007, p. 92).

Reception

The importance of the reception of audiovisual material has been emphasised by authors such as Mayoral (2005, p. 4), who believes that paying attention to the product may shed some light on the process, particularly when it is the addressee's reception of the product what is being examined. According to Miquel (2014, p. 64), research on reception is required both on a micro-level – to find out viewers' needs and preferences in order to arrive at a suitable translation solution – and on a macro-level, to gain empirical evidence of how spectators deal with the semiotics of the AVT products.

Di Giovanni (2016, p. 61) describes media reception studies as 'the empirical study of the social production of meaning in people's encounter with media discourses'. In this sense, Fuentes Luque's (2001) research on the reception of humour is one of the pioneering works investigating audience reaction in the field of AVT.

Other studies have gathered the opinion of spectators by using a scientific methodology. For instance, De los Reyes (2015, pp. 1-2) uses dubbed clips and a questionnaire to measure the reaction of spectators; Bourne and Lachat (2010, pp. 316-317) focused on audiodescription whereas the work of Antonini (2005, p. 216) followed a similar design but covering subtitled humour. Other prominent works in the field include several authors from the University of Bologna (De los Reyes, 2015, pp. 57-61), such as Rachele Antonini, but also Delia Chiaro, Chiara Bucaria and Linda Rossato. The work of Lång, Mäkisalo, Gowases, and Pietinen (2013, p. 77) in which subjects are observed with an eye-tracker while they evaluate subtitles, is also inspiring.

All in all, there is still a need for further research on reception studies of audiovisual material (Antonini, 2005, p. 210; Bucaria & Chiaro, 2007, p. 94; Bourne & Lachat, 2010, p. 335; Jekat et al., 2015, 448; Fuentes Luque, 2005, p. 140). There are only a few studies on acceptability in the AVT, therefore there is a call for further empirical research above all (Chaume, 2004, pp. 14-20; Chaume, 2007, p. 78). Besides, little progress has been made on this field and it tends to address the same aspects (De los Reyes, 2015, p. 36). Thence, 'gathering and evaluating the opinions of receivers of translated films undoubtedly has great potential' (Di Giovanni, 2012, p. 496).

MATERIALS AND METHODS

Most studies describe synchronies not only as characteristic features of AVT, but also as one of the relevant points to measure its quality. Furthermore, it is said that working without the help of the image reduces the quality of the products; as a consequence, it might be expected that the lack of the image affects the evaluation of the audiovisual products as well.

The aim of this experiment is to look for trends related to the perception of the quality of translation and synchrony by means of an analysis of the direct evaluation of various excerpts presented, either on

Dischronies and the Absence of Image in the Evaluation of Dubbing

video or as transcriptions in text format. The methodology is designed to confirm or refute the following hypotheses:

Hypothesis One: The excerpts with a lack of synchrony are given lower grades than the control ones.

Hypothesis Two: The lack of I shows the greatest influence in the rating given, as suggested by Chaume (2012, p. 68; 2005, p. 8).

Hypothesis Three: The subjects with a higher level of translation expertise perceive the lack of synchrony more clearly, thanks to their experience in translation and in AVT in particular.

Hypothesis Four: Seeing the image with the text has an effect on the grade given.

A selection of excerpts containing examples of different types of dischronies (I, L and K) and control scenes with no dischronies of any kind were used to carry out the experiment. The first step to carry out the experiment was to delve into relevant literature in order to check whether some scholars had previously mentioned paradigmatic examples of dischronies. The search was unfruitful; while many authors provide a definition and overview of I, L or K, there is very limited work offering specific examples about the lack of such synchronies (see Future Research Directions). Therefore, it was difficult to find excerpts that were suitable for the experiment³ (i.e. excerpts where different types of dischronies were found) and the search was based on the intuition of the researcher who had to look for help in other studies on the field of AVT. A sample of eight scenes was used to conduct the experiment two with L errors; two with K errors; two with absences of I and another two to be used as control scenes. Table 1 presents the scenes used.

The excerpts of the eight scenes chosen are provided in Appendix. Various and diverse examples were used: B-movies alongside well-known ones; television productions with others conceived to be released at the cinema; British and American works, etc. The control excerpts were selected in order to

Table 1. Used scenes

Code	Work	Data	Type	Description of the Scene
ACAD	<i>Combat Academy</i>	American television film (1986)	I	The dubbing outstretches the actor's mouth movements.
BRIA	<i>Monty Python's Life of Brian</i>	British film (1979)	K	The text of the dubbed version does not match the actor's hand gesture.
CAND	<i>I Want Candy</i>	British film (2007)	L	The dubbed version lasts much longer than the original one (Martínez Sierra, 2009)
NESS	<i>The Untouchables</i>	American film (1987)	I	The character of Ness moves his lips, but no sound is heard in the dubbed version.
PULP	<i>Pulp Fiction</i>	American film (1994)	K	The square drawn in the air completes an expression understood in English, but not in Spanish, therefore, an explanation would be required.
SKIN	<i>Skins</i>	Episode 1 of the 1 st season, British series (2007)	(Control)	Good translation in which the actor's movements have been taken into account.
TOXI	<i>The Toxic Avenger</i>	American B-movie (1984)	L	In the close-up ⁴ of the girl, the dubbed version does not match the movement of her lips.
TROL	<i>Troll 2</i>	Italian-American B-movie (1990)	(Control)	Another close-up, but this time there is a good lip adjustment.

offer good solutions to possible synchrony problems (in the case of SKIN, related to K; in the case of TROL, related to L).

The text of each scene was then transcribed and was provided to participants together with a brief description of the situation of each scene. In some cases, the author searched for the scripts of the works on the Internet, but transcripts were always checked against the exact version heard on the video to ensure both versions were identical. As a matter of fact, on a couple of occasions, the original version on video differed from the one published as a script on the Internet, which suggests that some of the scripts found on the Internet were not final versions.

In order to avoid the existing order effects when scenes are serially assessed (Muñoz & Conde, 2007, pp. 443-444), these were disposed randomly and differently among the subjects, thanks to a random number generator. Moreover, the subjects did not evaluate the same excerpt twice (firstly on video and secondly on text, or vice versa) in order to avoid ‘learning’ or evaluating with the previous version in mind. As a result, each subject evaluated the eight excerpts, but four of them were presented to subjects on video and the other four, in text format. Hence, all the files were evaluated the same number of times (i.e. 16).

Each subject randomly received a numerical code and a sequence, which they had to evaluate in the order given; for example:

Subject 209 received the sequence EPKGLCNM. The letters corresponded to: PULP text, SKIN text, ACAD video, TROL video, CAND text, NESS text, TOXI video and BRIA video;

Subject 404 received the sequence IKJPCGFH: TOXI text, ACAD video, CAND video, SKIN text, NESS text, TROL video, BRIA text and PULP video;

and so on. Letters were assigned to excerpts randomly, as Table 2 shows.

The excerpts were evaluated by students of the degree in Translation and Interpreting at the University of the Basque Country. In particular, two groups of 16 students were considered: second-year undergraduate students (hereafter 2U) and fourth-year undergraduate students (4U). A post-task questionnaire was included, from which the average profile of the subject was drawn, as well as their perceptions about AVT and the quality of the translation.

Table 2. Excerpts’ letters

Code	Letter According to the Presentation of the Excerpt	
	Text	Video
ACAD	O	K
BRIA	F	M
CAND	L	J
NESS	C	B
PULP	E	H
SKIN	P	D
TOXI	I	N
TROL	A	G

Dischronies and the Absence of Image in the Evaluation of Dubbing

The average subject is a Spanish native female enrolled in the degree in Translation and Interpreting with a C1 or C2 level in English, and is used to watching dubbed films, but not reading the script transcriptions. Most of the 2U are 19 years old and the 4U are 21. The average subject has no experience as an audiovisual translator (excluding what 4U have learnt in class). Eighty-one% of the 2U and 94% of the 4U affirm that they would like to find a job as translator; on the other hand, 31% of the 2U and 50% of the 4U would like to have a job as audiovisual translator in particular.

When designing the study, the researcher decided that the experiment would be conducted with students, since they are usually willing to participate in research studies voluntarily. This would also allow to include two different study levels, with different level of educational expertise, as part of the experiment. The group of 2U had not learnt about the challenges of AVT in class. However, the group of 4U had received more translation training during their degree and had just received concrete explanations about AVT and the importance of synchrony as part of a dedicated module on AVT. Consequently, it was expected that there would be differences in the data between the two subgroups of subjects. Another advantage of using students as subjects is that the experiment could be considered as a diagnostic evaluation tool, i.e. the test could reveal the knowledge the 2U and 4U had about the importance of synchronies and the support of the image for AVT, which could help the teachers plan the content for the classes during the rest of the course and subsequent academic years.

Once the test was designed, the researcher asked the teachers of the working groups to allow him to attend their classes to introduce the experiment to the students, who were informed they were going to take part in a study on the quality of dubbing and synchronies. The procedure used for both working groups (2U and 4U) was the same: the students were asked to download a compressed file that contained the instructions of the experiment, an evaluation sheet, a questionnaire and 16 folders. Each folder was named with a letter between A and P and included one of the excerpts, either on video or in text format (see Table 2). The subjects were asked to open the file with the instructions and follow these.

The researcher randomly distributed the numerical codes and sequences of letters that each subject had to evaluate. Students wrote the eight letters in the corresponding order in the evaluation sheet, and they opened the folders with the excerpts they had to evaluate one by one and following the corresponding order. After watching the scenes or reading the transcriptions, the subjects had to assign a rating between 1 and 4 (1 was the minimum grade and 4, the maximum) and provide a justification. As in similar research studies (e.g. Conde, 2009, p. 271) a four-level was used to avoid the tendency of participants placing most of the assessable elements in an assumed intermediate level.

When the subjects finished evaluating the eight excerpts, they saved the sheet with the assigned numerical code and answered the anonymous questionnaire (students were asked to write their corresponding code to preserve anonymity). The use of a numerical code made the analysis and presentation of the data easier and allowed to establish links between each subject's profile and their performance in the test. Once all data were saved, the subjects sent both files to the researcher by e-mail. Subjects took approximately one hour to complete the whole task.

The data analysis was carried out with the help of the SPSS v. 22 and MS Excel software.

SOLUTIONS AND RECOMMENDATIONS

This section presents results concerning the lack of synchrony with an emphasis on the role that the image plays for the evaluation of quality. The results are firstly discussed in general terms, followed by a comparison and discussion of the results of both working groups (2U and 4U).

How the Lack of Synchrony Affects Evaluation of Excerpts

Table 3 shows the average qualification obtained by each excerpt, with the maximum rating being 4 and the minimum being 1.

As the data shows, the highest-rated scene was the video version of *Skins*, which was one of the excerpts introduced as control scene. The text format of this scene also achieves a high rating; it seems that the evaluators kept in mind that translators had successfully solved a complicated problem a priori, since one of the actor's movements had to be considered – when he grabs other character's testicles. This claim is supported by subject 415, who justifies the high rating awarded to this version by stating that the image was very useful to understand the text and that both image and text were linked to each other.⁵ This high rating could also be due to the content of the series; *Skins* is a show for young people and the evaluators were young as well, so it could be the case that they identified themselves with the storyline.

The second highly rated work was *Troll 2*, which has little to do thematically with *Skins* but was the second control scene. Therefore, the two scenes without synchrony problems were the ones most highly rated by students. However, to test whether or not the differences between the control scenes and the other ones were significant, a statistical procedure was used (independent samples t-test). For the test, (control mean = 3.50, sd = 0.23, N = 4; no control mean = 3.04, sd = 0.41, N = 12) the 95% CI for the difference in means is -0.01, 0.93 ($t = 2.05$, $p = 0.059$, $df = 14$); hence, the null hypothesis that the true difference in means was zero could not be rejected, so the fact that an excerpt was a control scene did not affect the rating issued by the evaluators.

What is surprising is that the rest of the works were arranged depending on the type of synchrony issue, as Figure 1 shows.

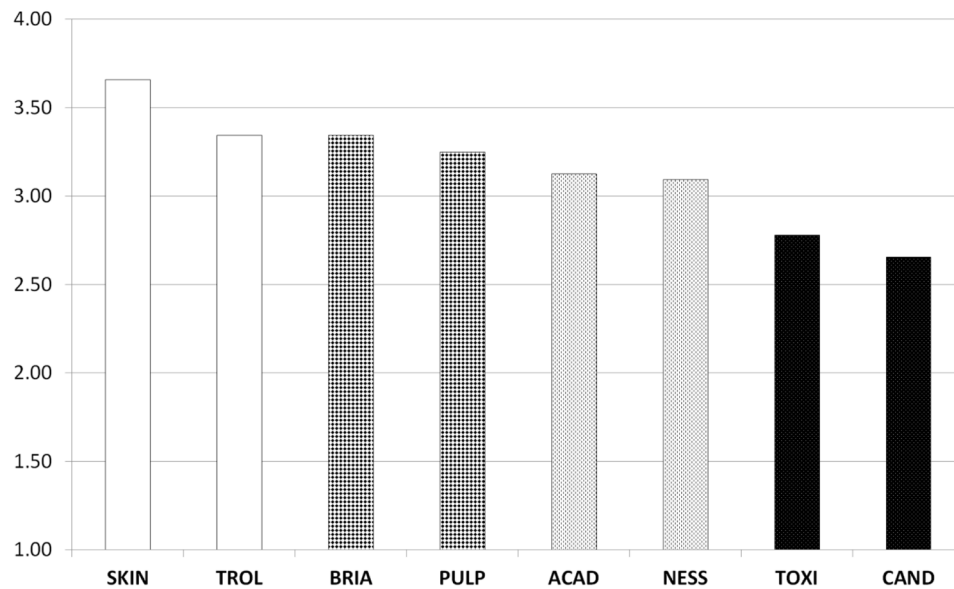
The results show that the highest ratings after the control scenes (white bars) were those showing absences of K (pattern filled with squares); followed by scenes with I problems (with similar ratings),

Table 3. Average rating per scene

	Text	Video	Av
ACAD	2.813	3.438	3.126
BRIA	3.250	3.438	3.344
CAND	2.063	3.250	2.657
NESS	2.813	3.375	3.094
PULP	3.063	3.438	3.251
SKIN	3.563	3.750	3.657
TOXI	2.625	2.938	2.782
TROL	3.188	3.500	3.344

Dischronies and the Absence of Image in the Evaluation of Dubbing

Figure 1. Average rating per work



and finally, the two examples with L errors (black bars). Table 4 shows a breakdown of data per group of subjects.

Both groups seem to have a similar perception of quality in general terms since the total average ratings assigned is alike (2U, 3.14 and 4U, 3.17). An independent-samples t-test was conducted to compare ratings issued by 2U and 4U. No significant difference was found in the ratings issued by 2U (mean = 3.14, sd = 0.45) and the ones issued by 4U (mean = 3.17, sd = 0.24); $t(22.79) = -0.24$, $p = 0.807$. These results confirm that the year of study does not seem have an effect on students' level of demand. In other words, having received AVT training in class or having more experience translating in general does not seem to affect the level of demand students show when evaluating the chosen samples.

Table 4. Average rating per scene and group of subjects

	2U			4U		
	Text	Video	Av	Text	Video	Av
ACAD	2.50	3.38	2.94	3.13	3.50	3.31
BRIA	3.38	3.50	3.44	3.13	3.38	3.25
CAND	2.00	3.00	2.50	2.13	3.50	2.81
NESS	2.88	3.25	3.06	2.75	3.50	3.13
PULP	3.13	3.50	3.31	3.00	3.38	3.19
SKIN	3.75	3.88	3.81	3.38	3.63	3.50
TOXI	2.50	2.75	2.63	2.75	3.13	2.94
TROL	3.25	3.63	3.44	3.13	3.38	3.25
Total	2.92	3.36	3.14	2.92	3.42	3.17

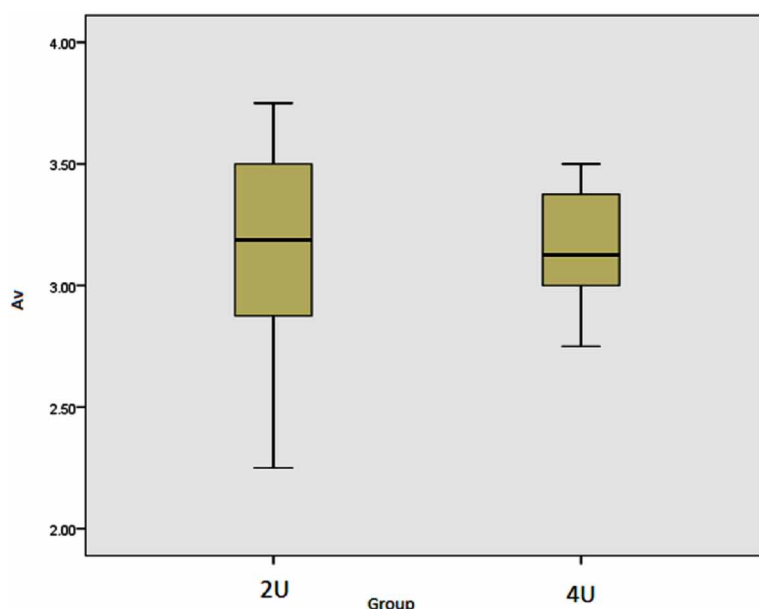
However, Figure 2 shows that even if averages are similar, 4U exhibit a more consistent behaviour compared to each other than 2U, therefore, a less pronounced standard deviation.

This could be due to the standardised effect carried out by education (Lachat, 2003); advanced students have a clearer view of the quality of the scenes and that is why their ratings are more homogeneous than those given by new students whose ratings are less consistent. Nevertheless, when a different independent samples t-test is undertaken per group, the results show that whereas for 4U the nature of the excerpt (control versus non-control) does not seem to affect the rating given (control mean 3.37, sd 0.20; no control mean 3.10, sd 0.40; $t = 1.73$, $p = 0.11$, $df = 11.01$, using Welch's procedure), there appears to be a slight effect between 2U (control mean 3.62, sd 0.27; no control mean 2.97, sd 0.47; $t = 3.37$, $p = 0.008$, $df = 9.48$, using Welch's procedure). In other words, when 2U work on a control excerpt, they tend to give a higher rating than when they have to assess a non-control excerpt. This does not happen amongst 4U.

The sequences containing L errors are the ones rated the lowest in both groups and the control scenes continue to be the best rated in the 2U group, but in the 4U group there is an excerpt (from ACAD), ranked between the two control scenes. Indeed, a great difference, present in the text version, is found in this excerpt between both groups; it seems that 2U did not approve the chosen equivalence while reading it (for example, subject 214 stated that the translator could have been more faithful to the original version), but their classmates approved it when watching the scene in video (student 201 said that even though the Spanish version had included some extra terms, these were important because the dubbed version was much more clear in that way). Nonetheless, 4U accepted both versions equally, maybe because they had learnt to distance themselves from the original work, offering more creative solutions.

Another interesting result is that 2U tend to evaluate the control excerpts more positively than 4U as well as those excerpts showing issues related to K dischronies; however, the result is the opposite with the other two types of issues (i.e. L and I). The results reveal that 4U assess the excerpts showing

Figure 2. Average rating per group



Dischronies and the Absence of Image in the Evaluation of Dubbing

errors related to L and I higher than 2U do. It could be expected that the advanced students would give more importance to these issues, which would, therefore, be reflected by a lower rating; however, this is not the case. Second-year undergraduate students may behave in a similar way to standard receivers, without the influence of the theory of translation and, thus, are more demanding. In contrast, 4U have received training on AVT and their judgment is not as influenced by the apparent dischronies, but by a more complex diversity of criteria. For instance, subject 412 mentions questions such as the spontaneity of language, or even the quality of the background noise to justify the rating provided.

The Impact of Image

The other main aspect of this research is the importance of the image for the evaluation of the scenes included in the experiment. It should be recalled that the students assessed the scenes either in dubbed video or directly on the transcription in text. Table 4 shows the average rating given to the excerpts depending on the format in which it was presented to the subjects.

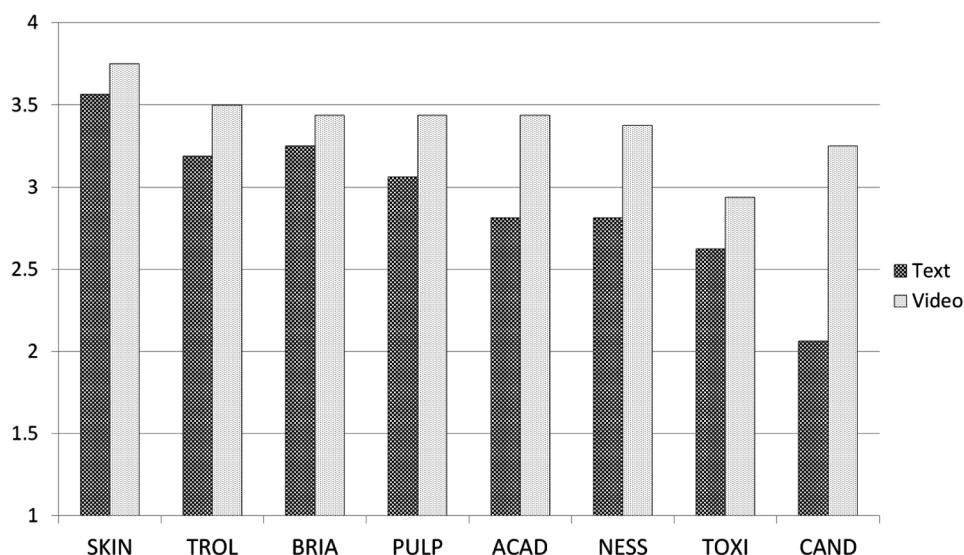
The scenes were generally rated more highly when evaluators could watch them on video than when scenes were only provided with the transcript. Again, an independent-samples t-test was calculated in order to check the results statistically, and Welch's procedure⁶ confirmed the initial results. There was a significant difference in the scores students awarded to the transcriptions (mean = 2.92, sd = 0.45) and the videos (mean = 3.39, sd = 0.23); $t(10.36) = 2.592, p = 0.026$. A translation that does not stand out at first by its quality, seem to gain with the image support – something that could be explained by the AVT's idiosyncrasy as intersemiotic experience and activity. As a matter of fact, individual data indicates this tendency is maintained regardless of whether the sample was chosen as an example of lack of synchrony or as control scene.

Furthermore, Figure 4 shows that not only were the video excerpts granted higher marks but also the ratings granted were far more alike than when the transcriptions were presented without the support of the image.

Two outliers appear among the excerpts presented in video: item 4, which corresponds to SKIN, and item 14, referring to TOXI. This means that, compared to the other excerpts assessed on video format, the dubbing of SKIN was extraordinarily highly evaluated; yet the dubbed version of TOXI had an average rating that was far from the one obtained by the scenes assessed with the help of the image. SKIN was a control scene whereas TOXI was an example of lip-synchrony. One could argue that excerpts with no problems of synchrony are well received while excerpts with lip dischronies are perceived as more problematic than others. Nevertheless, this only happens to one of the two examples of each excerpt type, and therefore, more examples should be tested and found to confirm this tendency.

At the same time, Figure 3, shows the rating gap between the video version and the text version is higher in some works than in others. The distance is particularly noticeable in CAND and in the two excerpts with problems related to I (i.e. ACAD and NESS). The comments are practically the same in both versions; that is, the subjects identify the errors both in video format and text format (besides the absence of I, they take into account the change between 'red lips' of the original and 'de grandes labios' of the Spanish translation). The lack or the excess of information is likely to become clear in the written text and might appear disguised through the synchrony on video thanks to the skills of the dubbing actors (that accelerate or reduce the rhythm of elocution), or to the existence of the image itself, which attracts the attention of the subject and distracts from the linguistic content. All in all, the synchrony would be

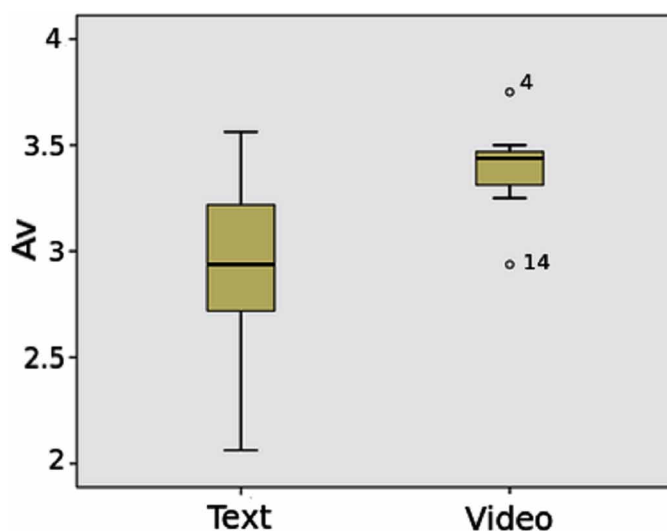
Figure 3. The impact of the image in the rating, per work



a sort of compensation technique in which the image or the elocution of the dubbing actors are the ones who fix or disguise the patent error in the text version. Figure 5 shows a breakdown of data per groups.

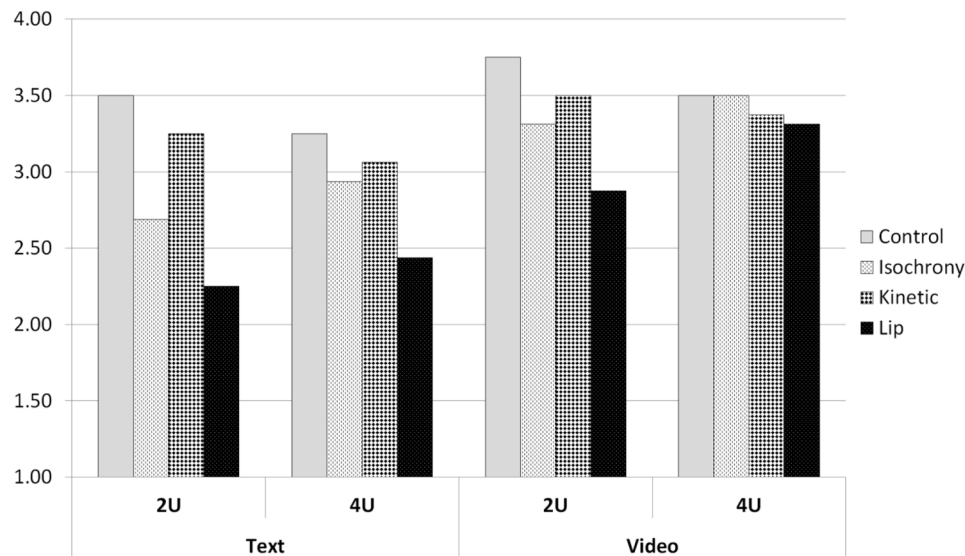
The pattern is similar in the four sets of columns; the scenes that were presented in video obtained higher scores. Besides, control scenes and those with some kind of K dischrony tend to get higher ratings, the exception being those video excerpts evaluated by 4U, which show very consistent numbers. This result is surprising, since 4U were expected to have a higher level of expertise to detect dischronies on video.⁷ It could be due to the fact that 4U are learning to translate audiovisual works; they are able to put themselves in the translators' shoes as they know how hard it is to adjust the text to the image, and

Figure 4. Average rating depending on the presentation format



Dischronies and the Absence of Image in the Evaluation of Dubbing

Figure 5. The impact of the image in the rating, per group



that empathy may push them to rate the samples more generously. In contrast, 2U would be stricter and therefore assign lower ratings because they lack that empathy.

A non-parametric test (i.e. the Mann-Whitney U test) was carried out. The results showed that there were no significant differences between groups either for text presentation or for video presentation. For text, the mean range among 2U (2.92) and 4U (2.92) is exactly the same; U is 123.5 and the associated p -value = 0.867. For video, the mean range among 2U (3.35) and 4U (3.42) is very similar as well, U is 117 and the associated p -value = 0.696. This means that Hypothesis Four cannot be confirmed and the distinct feature found among the video assessments carried out by 4U could have another explanation and, in any case, cannot be considered to be as significant as it appeared when looking just at the figures.

FUTURE RESEARCH DIRECTIONS

In light of the data, a reflection about what is really assessed when evaluating dubbing would be necessary, because it seems that synchrony errors do not have such a significant effect: 'The notion of subtitling quality deserves a more detailed study, as it seems possible that it consists of something entirely different from what has been thought before' (Lång et al., 2013, p. 84). It would be interesting, for example, to analyse whether the quality or nature of the original work affects the evaluation of its dubbing, something that has not been proven with the experiment carried out for this study (the excerpts taken from B-movies or television films were not assigned significantly higher or lower grades), probably within the framework of a larger study than the present one. The same applies to the discussion of the atypical values, for example, the excerpt with L errors that received atypically low grades (see Figure 4) when assessed with the help of the video.

As a consequence, further efforts are required to study these results empirically, as well as the idiosyncrasy of AVT in general: 'More research is still needed on how visual and verbal information are

actually processed and how they complement each other' (Miquel, 2014: p. 74). This project should be used as an inspiration for others where, for example, results might be compared with others derived from surveys or interviews with enterprises, production companies, professionals, teachers and other kinds of addressees who are not familiar with the theory and practice of AVT. In addition, some studies could be carried out about cases as diverse as the inherent benefits of relying on the video to perform the same activity of translating audiovisual texts. In this regard, more attention should be paid to another preliminary result: the fact that the excerpts assessed with the help of the video received more matching ratings than those presented in text format.

Furthermore, the study of translation expertise should be boosted within AVT research, for it is bringing interesting findings in other areas within translation studies (see e.g. research carried out by PETRA research group). Other groups of subjects should be tested to investigate how expertise influences the way AVT products are received and assessed. The time has come to link theory and practice and to show that experts not only translate texts differently from non-experts but also have other expectations relating to the quality of products and processes. Di Giovanni (2016, p. 76) compared the results of three independent experiments that measured the reception of subtitles and did not find any significant differences between those including experts (in cinema) and those including non-experts (general audience). However, they were different experiments, with diverse materials and methods and the results were compared only a posteriori; that is why it could hardly be said that this study is based on the concept of translation expertise.

Finally, it should be stressed that it was difficult to find examples of dischronies or specific problems in the related literature. A good starting point for future research about the incidence of synchronic errors in the reception of audiovisual texts would be perhaps the creation of a database where examples of dischronies could be classified and analysed previously by experts. This is not a light task, for when it comes to dubbing, at least in Spain, it appears to be of really high quality, though it might be worthwhile, given the position occupied by this concept in the literature on AVT.

CONCLUSION

Many renowned authors in the field of AVT call for the need of working with a video copy of the original. The present chapter investigates whether such a need might be addressed when evaluating dubbed translations. To that end, the starting point was an essential feature in dubbing: synchrony. As checked in the literature review, synchrony is thought to be in fact one of the main aspects to explain the quality of a dubbed video.

An experiment with translation students enrolled in different levels of study and with different level of expertise (i.e. 2U and 4U) was designed to check potential differences owing to the training received. This variable makes it possible to link AVT and dubbing with the concept of expertise, a field of research that possibly has a promising future within translation studies. In the research design, ethical issues were taken into account; particular attention was given to ensure anonymity as well as a simple experiment in terms of design, so that subjects could perform a common action (i.e. watching films and TV series) without requiring a great effort on their part.

Hypothesis One was about the relationship between quality perceived and whether the excerpts assessed introduced synchrony errors or not. The control scenes with no synchrony problems were given

Dischronies and the Absence of Image in the Evaluation of Dubbing

higher scores than those presenting some issues in this regard; however, the difference was not statistically significant. In any case, a noticeable gap was found within the 2U group; that is, they did perceive that an excerpt was a control scene and awarded higher ratings.

Hypothesis Two aimed to confirm if the lack of I had the greatest influence in the rating given by the subjects. In principle, the lowest-evaluated scenes were not the ones with I errors, as was expected according to the literature, but those with L errors, although the differences found were minor.

As for Hypothesis Three, 4U did not perceive the issues with synchrony or, at least, this lack of synchrony did not lead to lower ratings than those given by 2U. On the basis of the justifications that they introduced in the evaluation sheet, it was found that some 2U took into account the synchrony errors in their ratings, which, contrary to what was expected, they were very aware of, perhaps due to personal or vocational preferences. On the other hand, excerpts containing L and I errors were given higher grades by 4U, and it has been suggested that perhaps this group of subjects does not necessarily evaluate scenes on the basis of synchrony but also on other criteria that they are more familiar with, due to their level of expertise. The ratings of 4U were more consistent within the group, which suggests that they were perhaps more confident about what a translation of quality is in AVT.

Finally, contrary to expectations, and regarding Hypothesis Four, having the video file to evaluate the dubbing quality did not change the subjects' level of demand on the basis of the synchrony errors they detected. When students watched the excerpt, they rated the dubbing higher, possibly because more elements came into play in order to make a judgment value or because the errors that were clear in the text were counterbalanced by the dubbing actors or the final cut.

The results reveal the importance of dubbing as an activity far beyond the linguistic level, where different aspects, such as the tone, the actors' elocution or even the soundtrack of the works might becloud or disguise possible errors made when translating or adjusting. The present research, although limited in its scope, has served to show the predisposition of the students towards AVT and the knowledge, more or less intuitive, that they had about synchrony.

As stated in the Materials and Methods section, the experiment has proven useful as a diagnostic test. On the one hand, 2U tend to be more impressed by the existence of errors related to synchrony and that is possibly the reason why they award higher grades to control excerpts. In a way, they behave as a standard, non-expert audience. On the other hand, the evaluations given by 4U are more consistent but, at the same time, are more lenient towards the excerpts including synchrony errors. This is perhaps because they have developed a sense of empathy with the translators and because they have learnt that the quality of AVT does not depend merely on synchrony but also on other questions such as the translators' creativity or the recording's quality, among others.

Translation teachers could draw conclusions from the results concerning the two groups who took part in the research. Second-year undergraduate students could learn that quality in AVT is not only a question of precise or defective synchrony. In addition, 4U should be encouraged to reflect about their role as semi-experts and how their level of expertise could influence the way they receive and assess audiovisual products; in other words, the fact that there is a distance between their own expectations and those of the typical audience. Finally, teachers could initiate a debate on the importance of errors and appropriate solutions in AVT, once the students have learnt to include different criteria in their evaluation, they need a way to establish a hierarchy among those criteria.

ACKNOWLEDGMENT

This research was supported by the University of the Basque Country, UPV/EHU. TRALIMA Consolidated Research Group GIU16_48. Project FFI2012-39012-C04-01T and G15/75 MINECO, Spanish Ministry of Economy, Industry and Competitiveness.

REFERENCES

- Abdallah, K. (2012). *Translators in production networks. Reflections on agency, quality and ethics*. Joensuu, Finland: Publications of the University of Eastern Finland.
- Agost, R. (1999). *Traducción y doblaje: Palabras, voces e imágenes*. Barcelona, Spain: Ariel.
- Alfaro, C. (2012). Quality standards or censorship? Language control policies in cable TV subtitles in Brazil. *Meta: journal des traducteurs/Meta: Translators' Journal*, 57(2), 464–477.
- Antonini, R. (2005). The perception of subtitled humour in Italy. *Humor: International Journal of Humor Research*, 18(2), 209–225. doi:10.1515/humr.2005.18.2.209
- Arumí, M., Carrabina, J., Matamala, A., Mesa-Lao, B., Orero, P., & Serrano, J. (2013). Audiovisual translation learning platform: A new tool for audiovisual translation training. *Hermeneus*, 15, 39–66.
- Bittner, H. (2011). The quality of translation in subtitling. *Trans-kom*, 4(1), 76–87.
- Bourne, J., & Lachat, C. (2010). Impacto de la norma AENOR: Valoración del usuario. In C. Jiménez, A. Rodríguez, & C. Seibel (Eds.), *Un corpus de cine. Teoría y práctica de la audiodescripción* (pp. 315–333). Granada, Spain: Tragacanto.
- Bucaria, C., & Chiaro, D. (2007). End-user perception of screen translation: The case of Italian dubbing. *TradTerm*, 13, 91–118. doi:10.11606/issn.2317-9511.tradterm.2007.47468
- Chaume, F. (2004). *Cine y traducción*. Madrid, Spain: Cátedra.
- Chaume, F. (2005). Los estándares de calidad y la recepción de la traducción audiovisual. *Puentes*, 6, 5–12.
- Chaume, F. (2007). Quality standards in dubbing: A proposal. *TradTerm*, 13, 71–87. doi:10.11606/issn.2317-9511.tradterm.2007.47466
- Chaume, F. (2012). *Audiovisual translation. Dubbing*. Manchester, UK: St Jerome.
- Conde. (2009). *Proceso y resultado de la evaluación de traducciones* (Unpublished doctoral dissertation). University of Granada, Granada, Spain. Retrieved October 27, 2017, from <http://www.tdx.cat/handle/10803/16836>
- De los Reyes, J. (2015). *La traducción del cine para niños. Un estudio sobre recepción* (Unpublished doctoral dissertation). Jaume I University, Castellón de la Plana, Spain. Retrieved October 27, 2017, from <http://www.tdx.cat/handle/10803/334991>

Dischronies and the Absence of Image in the Evaluation of Dubbing

De Yzaguirre, L., García, Y., Matamala Ripoll, A., & Ribas, M. (2005). Tratamiento informático de un corpus de guiones y traducciones de textos audiovisuales: Implicaciones en traductología. In P. Zabalbeascoa, L. Santamaría, & F. Chaume (Eds.), *La traducción audiovisual: Investigación, enseñanza y profesión* (pp. 107–113). Granada, Spain: Comares.

Di Giovanni, E. (2012). Why not ask the audience? An attempt to evaluate viewers' reception of subtitled films at festivals. In F. Dalziel, S. Gesuato, & M. T. Musachio (Eds.), *A lifetime of English studies. Essays in honour of Carol Taylor Torsello* (pp. 471–486). Padua, Italy: Il Poligrafo.

Di Giovanni, E. (2012). Why not ask the audience? An attempt to evaluate viewers' reception of subtitled films at festivals. In F. Dalziel, S. Gesuato, & M. T. Musachio (Eds.), *A lifetime of English studies. Essays in honour of Carol Taylor Torsello* (pp. 471–486). Padua, Italy: Il Poligrafo.

Di Giovanni, E. (2016). Reception studies in audiovisual translation research. The case of subtitling at film festivals. *Trans-kom*, 9(1), 58–78.

Díaz-Cintas, J. (2003). *Teoría y práctica de la subtitulación inglés-español*. Barcelona, Spain: Ariel.

Fuentes Luque, A. (2001). *La recepción del humor audiovisual traducido: Estudio comparativo de fragmentos de las versiones doblada y subtitulada al español de la película «Duck Soup» de los Hermanos Marx* (Unpublished doctoral dissertation). University of Granada, Granada, Spain. Retrieved October 27, 2017, from <http://www.tdx.cat/handle/10803/145573>

Fuentes Luque, A. (2005). La recepción de la traducción audiovisual: Humor y cultura. In P. Zabalbeascoa, L. Santamaría, & F. Chaume (Eds.), *La traducción audiovisual: Investigación, enseñanza y profesión* (pp. 139–152). Granada, Spain: Comares.

Hurtado, A. (2001). *Traducción y traductología. Introducción a la traductología* (6th ed.). Madrid, Spain: Cátedra.

Jekat, S. J., Prontera, D., & Bale, R. (2015). On the perception of audio description: Developing a model to compare films and their audio described versions. *Trans-kom*, 8(2), 446–464.

Lachat, C. (2003). *Problemas y estrategias de traducción* (Unpublished doctoral dissertation). University of Granada, Granada, Spain. Retrieved October 27, 2017, from <http://www.tdx.cat/handle/10803/18013>

Lång, J., Mäkisalo, J., Gowases, T., & Pietinen, S. (2013). Using eye tracking to study the effect of badly synchronized subtitles on the gaze paths of television viewers. *New Voices in Translation Studies*, 10, 72–86.

Martínez Sierra, J. J. (2009). The challenge of translating humour for dubbing. Some problematical issues. In A. Goldstein & B. Golubović (Eds.), *Foreign language movies – Dubbing vs. subtitling* (pp. 129–149). Hamburg, Germany: Verlag Dr. Kovač.

Mayoral, R. (2001). El espectador y la traducción audiovisual. In F. Chaume & R. Agost (Eds.), *La traducción en los medios audiovisuales* (pp. 33–46). Castellón de la Plana, Spain: Publicacions de la Universitat Jaume I.

Mayoral, R. (2005). Reflexiones sobre la investigación en traducción audiovisual. In P. Zabalbeascoa, L. Santamaría, & F. Chaume (Eds.), *La traducción audiovisual: Investigación, enseñanza y profesión* (pp. 3–8). Granada, Spain: Comares.

- Miquel, M. (2014). The reception of subtitling by the deaf and hard of hearing. Preliminary findings. In E. Torres-Simon & D. Orrego-Carmona (Eds.), *From translation research projects 5* (pp. 63–76). Tarragona, Spain: Intercultural Studies Group.
- Muñoz, R., & Conde, T. (2007). Effects of serial translation evaluation. In P. A. Schmidt & H. E. Jüngst (Eds.), *Translationsqualität* (pp. 428–444). Frankfurt, Germany: Peter Lang.
- Orero, P. (2005). La traducción de entrevistas para voice-over. In P. Zabalbeascoa, L. Santamaría, & F. Chaume (Eds.), *La traducción audiovisual: Investigación, enseñanza y profesión* (pp. 213–222). Granada, Spain: Comares.
- Segovia, R. (2005). Semiótica textual y traducción audiovisual. In P. Zabalbeascoa, L. Santamaría, & F. Chaume (Eds.), *La traducción audiovisual: Investigación, enseñanza y profesión* (pp. 81–92). Granada, Spain: Comares.
- Sokoli, S. (2005). Temas de investigación en traducción audiovisual: La definición del texto audiovisual. In P. Zabalbeascoa, L. Santamaría, & F. Chaume (Eds.), *La traducción audiovisual: Investigación, enseñanza y profesión* (pp. 177–185). Granada, Spain: Comares.
- Waddington, C. (2001). Should translations be assessed holistically or through error analysis? *Hermes*, 26, 15–38.
- Weidmann, A. (2014). Critiquez-nous! Mais critiquez-nous bien. Pour une critique grand public des traductions/adaptations audiovisuelles. *L'Écran traduit*, 3, 45-71.

KEY TERMS AND DEFINITIONS

Audiovisual Translation: Within translation studies, a modality that deals with the transfer of multimodal and multimedial texts (i.e., those that transfer information through both acoustic and visual channels and convey meaning by the interaction of various codes). There are many forms of audiovisual translation, with the main ones being dubbing and subtitling.

Dubbing: One of the most important forms of audiovisual translation. It consists of the re-recording of the original voice tracks in the target language in the voices of the dubbing actors. It is by far the most widespread way of translating audiovisual products in Spain, Germany, Italy, and France, among others.

Excerpt of Scene Transcription: In this chapter, it refers to the transcription of a dubbed scene or excerpt as it is perceived by the researcher. It consists of a brief description of the situation, followed by the sentences uttered by the actors that appear on screen (both in the English original version and in the Spanish dubbed version).

Isochrony: In dubbing, it refers to a form of visual synchrony that adjusts utterances and pauses of the original work with the utterances and pauses introduced in the dubbed version. Some authors consider it the most significant type of adjustment, for it is thought to be easily detected by viewers.

Kinetic Synchrony: In dubbing, it refers to another form of visual synchrony that measures how the information uttered by the voice delivering the translation coincides with the information that is transmitted by the movements of the actors or their bodies as they appear on-screen. A paradigmatic example of a lack of kinetic synchrony in Western countries is that of a shaking head expressing negation accompanied by the word “yes.”

Dischronies and the Absence of Image in the Evaluation of Dubbing

Lip Synchrony: Also in dubbing, it refers to a form of visual synchrony that implies the correspondence of the screen character's movements of the lips and the translated version. It is said to be especially relevant in close-up shots.

Order Effects: This concept is used in different fields of psychology and linguistics and refers to the differences in the responses given by the research subjects that may be explained by the order in which the materials are presented to them. In translation studies, it is usually taken into account in works inspired by cognitive translatology and translation process research.

Reception Studies: Works and investigations concerning the audience's response to audiovisual products and other forms of communication. This approach comes from the literary theory of reader response; nowadays, it is also investigated in audiovisual translation.

Translation Evaluation: The process of assigning a qualification or mark to a given translation based on its quality and usually after an analysis has been carried out. There is no unique way to evaluate translations for it depends on the translation itself and the goal of the evaluation. Sometimes (and it is also the case in this chapter) the term *assessment* is used as a synonym for *evaluation*, but some authors establish a difference between both concepts.

Translation Expertise: The skills and capabilities of expert translators, i.e. those translators who have an extensive record in their domain and thus perform their work with exceptional results when compared to other colleagues with less experience. Research based on this approach often compares the performance of experts and non-experts or novices. In this chapter, *translation expertise* also includes expertise acquired within the educational framework.

ENDNOTES

- ¹ This definition concerns what some authors (Agost, 1999, pp. 58-59) call 'visual synchronization', i.e. one of the possible types of synchrony, and to which synchrony 'of characterisation' and 'of content' should be added. However, in this study the author follows Chaume's recommendation (2012, p. 70), i.e. to consider synchrony what Agost called 'visual synchronisation'.
- ² The congress 'Languages and the Media II', celebrated in Berlin in 1998, may be considered a turning point. According to Díaz-Cintas (2003, p. 70), in that congress a greater collective effort was proposed in order to achieve acceptable quality levels all over the world.
- ³ Arumí et al. (2013, p. 53), who used clips for their research, also found this task very time-consuming.
- ⁴ In cinema and television, it is a type of shot which frames tightly people or objects.
- ⁵ The subjects justified the ratings given in the evaluation sheet; as a result, some ideas offering numerical results can be explained to an extent.
- ⁶ Despite the fact that boxplots seem considerably different for both groups, Welch's procedure was used. The reason for this was that the *p*-value of the Levene's test was higher than 0.05 (0.109, indeed).
- ⁷ However, there are some precedents: in the study of Lång et al. (2013, p. 77) the evaluators did not perceive the errors introduced in the subtitles or, if they did, these errors did not affect the ratings given negatively.

APPENDIX

This section includes the excerpts of the scene transcriptions used in the experiment.

Table 5.

Troll 2			
<i>Close up of a boy with glasses (ARNOLD) who is looking upwards, scared.</i>			
ARNOLD:	They're eating her. And now they're going in me.	ARNOLD:	Se la están comiendo. Y ahora me van a comer a mí.
<i>The camera focuses on what ARNOLD is seeing: some little monsters are devouring a corpse. Now it focuses again on the boy's distorted face.</i>			
ARNOLD:	¡Oh, my God!	ARNOLD:	¡Oh, Dios mío!

Table 6.

Skins			
<i>A man with a moustache (DEALER) is sitting next to a young boy (SID) on a sofa.</i>			
DEALER:	Have you got balls?	DEALER:	¿Hay huevos?
SID:	Yeah.	SID:	Sí.
<i>The man puts the cigarette on his mouth and touches the private parts of the young boy. The boy groans in disgust.</i>			
DEALER:	I'm just checking, because your balls are your collateral.	DEALER:	Tenía que comprobarlo; los huevos serán tu aval.

Table 7.

Combat Academy			
<i>A man in military dress (WOODS) talking to other two: one with long hair (MAX) and the other with glasses.</i>			
WOODS:	We're here for supplies, that's all.	WOODS:	Hemos venido a comprar provisiones, nada más.
MAX:	Ah, not meaning the Hitler Youth, right?	MAX:	Ah, ¿no han venido a la reunión de juventudes hitlerianas?

Dischronies and the Absence of Image in the Evaluation of Dubbing

Table 8.

The Untouchables			
<i>NESS is being questioned by journalists.</i>			
NESS: JOURN. 1: JOURN. 2: JOURN. 3:	It is the law of the land. Do you consider yourself a crusader, Mr. Ness, is that it? What type of qualifications do you have for the job, Mr. Ness? Do you have any political ambitions?	NESS: JOURN. 1: JOURN. 2: JOURN. 3:	Es una ley del país. ¿Se considera usted una especie de cruzado, señor Ness? ¿Qué dotes especiales cree tener para este trabajo, señor Ness? ¿Tiene usted ambiciones políticas, señor Ness?
<i>Whispering. Flashes of photographic cameras. Ness walks through questions, protected by the police CHIEF.</i>			
CHIEF: NESS:	All right, boys. I think that's enough. Excuse me.	CHIEF:	Basta ya, señores. Por favor. Creo que ya es suficiente, por favor.
<i>While he walks among the crowd, another journalist approaches him, who asks him directly.</i>			
JOURN. 4: NESS:	What are your real plans? What have you got coming up? You'll just have to read about it in the paper.	JOURN. 4: NESS:	Dígame, ¿cuáles son sus planes?, ¿qué se propone? Ehm, tendrá que leerlo en los periódicos.

Table 9.

Pulp Fiction			
<i>A man (VINCENT) and a woman (MIA) sitting on the front seat of a car.</i>			
VINCENT: MIA:	Come on, Mia, let's go get a steak. You can get a steak here, daddy-o. Don't be a...	VINCENT: MIA:	Vayamos a comernos un filete. Puedes comerte uno aquí, colega. No me seas...
<i>Mia draws in the air a kind of square that seems to be superimposed on the image, thanks to the later editing of the film. The square disappears. Mia is about to get out the car.</i>			
VINCENT:	After you, kitty-cat.	VINCENT:	Tú primero, gatita.

Table 10.

Monty Python's Life of Brian			
<i>Brian and his mother (MANDY) walk among the people in the market while an EX-LEPROUS tries to haggle with them.</i>			
EX-LEP.: MANDY:	All right. Two shekels. Just two. Isn't this fun, eh? Look. He's not giving you any money, so piss off!	EX-LEP.: MANDY:	Bueno, lo dejamos en dos shekels. ¿A que es divertido? Oye. No te va a dar una perra, así que ¡fuera!

Table 11.

I Want Candy			
<i>Two boys talking on the phone. JOE walks down the street, while BAGLEY is sitting in front of a computer.</i>			
JOE:	And, tell me, what's she like?	JOE:	¿Qué? ¡Alucinante! Y... dime... ¿cómo es ella?
BAGLEY:	Well, she's got blonde hair, red lips, a nice... bum.	BAGLEY:	Bueno, pues... es rubia, de grandes labios, de culo... bien...
JOE:	Is she smart? You know, funny?	JOE:	¿Es lista? ¿Es simpática?
BAGLEY:	She's a doll.	BAGLEY:	Es una muñeca.
JOE:	Yeah. I hear you.	JOE:	Sí, claro. Y que lo digas.

Table 12.

The Toxic Avenger			
<i>A girl who cannot see (WANDA) has a monster in front of her (the TOXIC Avenger). She is sitting on the floor, crying and scared.</i>			
WANDA:	Could you take me out of here, please? Please. I'll never make it by myself without my dog. Please.	WANDA:	¿Puede sacarme de aquí, por favor? ¡Por favor! Es que yo sola no podré apañarme para salir de aquí...
TOXIC:	Okay, alright, but, come on, let's go out the back way. Hey, don't worry. I won't hurt you. Come on, I'll help you up.	TOXIC:	De acuerdo. Pero saldremos por la puerta trasera, ¿eh? Y no te preocupes, no te haré daño. Vamos.

Chapter 7

Quality Assessment in Audio Description: Lessons Learned From Interpreting

Louise Fryer
University College London, UK

ABSTRACT

Audio description (AD) is one of the younger modes of translation. It shares many similarities with interpreting, although AD users have specific needs because they are blind or partially sighted. As quality is of concern in both fields, this chapter explores the overlaps to see what can be learned for AD from research already carried out in interpreting. Macro and micro criteria suggested for each discipline are compared, and describer competencies are discussed in the context of AdlabPRO, a European research project that seeks to define the professional profile of an audio describer and develop training materials and courses. The chapter concludes that assessment protocols and rating scales developed for interpreting might be adopted for AD, after appropriate adaptation to accommodate areas where the fit is incomplete. These include synchrony and the need for the AD to be considered, not in isolation, but in relation to the existing audio elements of the source text (ST).

INTRODUCTION

Audio description (AD) is one of the access modes of audiovisual translation (AVT), a verbal commentary that makes visual information accessible to people who cannot perceive it unaided—principally people who are blind or partially sighted. Audio describers work in a range of settings, describing everything from recorded media (e.g., film and television) to live events such as theatre, opera, dance, and sport. Describers in live settings could be thought to perform a similar role to simultaneous interpreters (i.e., oral translation delivered live), and require many of the same competencies (Iglesias Fernández, Martínez Martínez, & Chica Núñez, 2015).

Interpreting as a discipline as well as a profession encompasses a broad range of activities (as cited in Pöchhacker, 2016). As Kohn and Kalina express it (1996, p.118), “The interpreter listens to the speaker’s

DOI: 10.4018/978-1-5225-5225-3.ch007

source discourse and produces a corresponding target discourse which will enable the target discourse audience to understand what the speaker meant.” For audiovisual texts, the audio describer watches the source product and produces a corresponding spoken discourse which will enable the target audience to understand the full message of the original. This chapter explores commonalities between AD and simultaneous interpreting, with particular regard to quality assurance and assessment. As Iglesias Fernández and her colleagues suggest (2015, p.75), “it seems only natural to investigate what quality assessment research into interpreting can teach us about quality reception in AD.” Although there are similarities between the quality of the end product (the description) and the skills and competencies of the translator (the describer), in that the former is dependent on the latter, this chapter argues that a distinction must be drawn between ways of assessing the quality of the description and assessing the skills and competencies of the describer, such that the describer has the skills needed to meet established standards. Pöschhacker (2001) points out that interpreting quality assessment for trainees is essentially different from professional assessment, because of its guiding function. This is discussed in the context of ADLAB PRO, a three-year research project financed by the European Union under the Erasmus+ Programme Key Action 2 – Strategic Partnerships. The project aims to define the profile of the professional audio describer, to identify current training routes and future training needs, and to create open-access training materials. The ADLAB PRO partners intend the resulting training course to be accredited using the European Credit Transfer System (ECTS) or The European Credit system for Vocational Education and Training (ECVETS), so that the quality of the course and its graduates may be recognised throughout Europe. One of the intellectual outputs of the project is a questionnaire (IO1, 2017a) that has provided a snapshot of the current training situation. Compiled by Agnieszka Chmiel and Iwona Mazur from the Adam Mickiewicz University in Poznan, Poland, the questionnaire sampled 192 extant AD courses across Europe, comprising 93 academic courses, and 99 non-academic courses. In this chapter, criteria for quality assessment in interpreter training will be compared with criteria for assessing trainee describers. The complexities of creating a numerical weighting system are discussed, and a solution proposed in the form of assessment sheets (contained in the appendix).

BACKGROUND

Gouadec (2010, p. 270) argues, “Quality in translation is both the quality of an end-product (the translated material) and the quality of the transaction (the service provided).” He acknowledges that only the quality of the service-provision transaction has been benchmarked through a number of standards, the best known being EN 15038— a quality standard specifically written for the translation industry. “The basic idea is that the quality of the transaction is ‘good’ if and when both the provider and providee are satisfied with the translation provision process and, of course, its result” (Gouadec 2010, p. 270). This fits happily with the marketing model that Ingrid Kurz (1989, 1993, 2001) has applied in relation to quality in conference interpreting, to which the discussion will turn below. Yet, as observed above, there is a third aspect of quality to consider: namely, the skills and performance of the trainee whose work needs to be assessed for pedagogical purposes.

Creating a course implies understanding the quality of the end product, so that the ingredients included in the course will produce learners capable of producing such quality. For a course to deliver high-quality training, those creating the course must understand the skills and competencies necessary to produce a quality product.

Quality Assessment in Audio Description

Most AD teachers sampled by the ADLAB PRO questionnaire were found to have experience in teaching and AD drafting. The implication is that they could draw on their practical experience while teaching and assessing. Survey results obtained by Torres-Simón and Pym (2016) from a questionnaire for translation scholars likewise found that “96% of them had translated or interpreted on a regular basis” (p. 110). This suggests that as expert practitioners, AD teachers are in a good position to understand what quality means for AD. Given that assumption, the ADLAB Pro questionnaire assembled a list of competencies from the literature and asked existing trainers/teachers to rank nine specific competencies on a five-point Likert scale (1=of no importance; 5=very important). The results are shown in Table 1.

MAIN FOCUS OF THE CHAPTER

While this chapter considers the multifaceted aspects of quality in AD it will focus on comparing competencies that have been suggested as indicators of quality for assessment purposes in both interpreting and audio description.

Issues, Controversies, Problems

In interpreting, doubt has been cast on the assumption that experts make the best judges of quality. For example, Kalina (2005) points out, “Interpreters and trainers feel that they can assess the quality of colleagues or trainees intuitively, on the basis of their experience and professionalism, but they are unable to express their subjective judgements by objectively measurable standards” (p. 768). Marzà Ibañez (2010), who has worked on quality and assessment in AD in the context of academic courses, agrees that professional describers develop an understanding of what is needed to create a quality AD through practice, but that this is not sufficiently clearly defined to enable students to develop the necessary skills. Marzà Ibañez (2010) turned to second-language assessment studies and existing AD guidelines to garner criteria, resulting in three main categories or macrocriteria: linguistic transfer, pragmatic and intersemiotic transfer, and time management. One difficulty is that AD guidelines in different countries are many and various, and even sometimes contradictory (Vercauteren, 2007). In order to overcome this, Cristóbal

Table 1. Index of the relative importance of AD competences

Competence	Importance (Out of 5)
Choosing the most relevant information to describe	4.87
Perfect use of mother tongue	4.70
Knowledge of the needs of the visually impaired	4.65
Choosing appropriate AD strategies (e.g. deciding when to name a character)	4.46
Technical aspects (editing, timing, text compression)	4.32
Knowledge of cinematography, theatre, arts and/or semiotics of the image	3.84
vocal skills	3.52
IT related skills (using appropriate software, etc.)	3.51
Reflecting filmic language (editing and camerawork) in AD	3.22

ADLAB PRO, 2017a.

Cabeza Cáceres (2017) took only noncontradictory items from various standards and came up with three macrocriteria: content, language, and sound. He subdivided them into the following microcriteria:

content>synchrony>no overlapping of AD with dialogue; **language**>articles>grammar>pronunciation>confusing phrasing
sound>voice talent>good balance of AD & soundtrack (level and fade) as shown in Table 2.

There are a number of issues with the model that Cabeza Cáceres proposes. First, by returning to guidelines that in some cases were created as long ago as 2000, he disregards reception research that has cast doubt on the validity of certain elements. Iglesias Fernández and her colleagues assert that “reception studies on AD are extremely scarce” (2015, p. 75). However, several published studies show that approaches enshrined by the guidelines are not always those favoured by AD users (e.g. Fryer & Freeman, 2012a, 2012b; Iglesias Fernández et al., 2015; Ramos, 2015; Walczak & Fryer, 2017a, 2017b). In addition, in terms of assessment of describers, Cabeza Cáceres includes elements that are out of the describer’s control, such as sound balance, although this aspect is clearly important for assessing the quality of the end product. Furthermore, he proposes a simple additive model based on error analysis, awarding one point for each transgression, and summing them to attain an overall score. Consequently, it is likely that the longer the duration of the AD product assessed, the higher the score would be. Under this accounting system, a higher score would automatically indicate worse quality. Finally, this type of negative model fails to reward any aspect of creativity on the describer’s part, such as choice of vocabulary.

For training conference interpreters, Schjoldager (1996, p.191) suggests four macrocriteria for giving feedback on students’ interpreting performance: comprehensibility and delivery; language; coherence and plausibility; and loyalty. Table 3 displays the quality criteria suggested for interpreting, compared with those currently proposed for AD. It is clear from Table 3, and from the categories identified by Cabeza Cáceres (2017), that competencies developed for interpreting and for AD have much in common.

Overlaps in Competencies

Comprehensibility

Comprehensibility would seem to be an obvious criterion of quality in any form of translation. The receiver must be able to understand the content if the translation is to succeed at all. Yet, it could be argued that this self-evident truth nonetheless depends on the *skopos*, or purpose, of the source text (ST). For example, in “cinema of attractions” (Matamala & Remael, 2015) —it may be more important that the

Table 2. Assessment macro and microcriteria proposed by Cabeza Cáceres

Content	Language	Sound
Synchrony	Articles	Voice talent
No overlapping of AD with dialogue	Grammar	Good balance of AD & soundtrack (level and fade)
	Pronunciation	
	Confusing phrasing	

Quality Assessment in Audio Description

Table 3. Assessment criteria proposed by interpreting Scholars (IS) and AD scholars

Schjoldager (IS)	Cristóbal Cabeza Cáceres(AD)	Pöchhacker (IS)	Lee (IS)	Marzà Ibañez (AD)
comprehensibility and delivery		successful communicative interaction	delivery	
language	language	adequate TL expression	TL quality	linguistic transfer
coherence and plausibility				
loyalty		accurate rendition equivalent intended effect	accuracy	pragmatic and intersemiotic transfer
	content			
	sound			
	synchrony			time management

AD enables its users to share an appreciation of the spectacle, rather than simply to understand what is happening. Certainly, AVT reception research, using *Presence* as a measure (see below), has consistently shown that comprehensibility and enjoyment are not related (Wissmath, Weibel, & Groner, 2009; Fryer & Freeman, 2014; Walczak & Fryer, 2017a, 2017b).

Comprehensibility does not appear in the list of competencies compiled by ADLAB PRO in the IO1 questionnaire. Cabeza Cáceres (2017) includes “confusion” which could be seen as the flipside of comprehensibility.

Delivery

Schjoldager (1996) couples comprehensibility with delivery. Not only does lack of vocal clarity (mumbling) affect comprehensibility, but also paralinguistic aspects such as segmentation and prosody (including intonation and pace) help to shape meaning (e.g., Sacks, 1998; Iglesias Fernández et al., 2015; Fryer, 2016). Delivery appears in the ADLAB PRO competencies under the banner of vocal skills, a competence ranked only eighth out of nine in importance. This may seem surprisingly low; however, it is worth remembering that practices differ in different modes of AD, from country to country and even within countries. In some settings (e.g., opera AD in Switzerland and France; screen AD in Poland, Italy, and Spain), professional voice talents are used, while in others (e.g., live AD and some screen AD in the UK), a describer will deliver his/her own script, making the emphasis on a competent delivery highly relevant. There are further geographical differences in the skills required, based on the language of the products to be described. For example, Orero (2005) cites a list of competencies put together by the UK Audio Description Association (ADA, 2004, p. 5) in its course manual, under the heading of “characteristics of a good describer,” including:

- “The ability to summarize information accurately and objectively;
- A good command of the language;
- A clear and pleasant speaking voice; Good sight and hearing (whether aided or unaided);
- The ability to work as part of a team; The commitment to access for disabled people,
- The provision of quality audio description to blind and partially sighted people.”

while noting that “these skills are drafted from the UK, a country with little import of media products in other languages than English. Other countries where there is a large import of audiovisual programmes, such as Spain, much of the accessible work will have to be done by a translator, or a person with foreign language transfer skills. It will also have to be decided who does the recording of the AD, since in Spain there is a dubbing union which protects the working rights and conditions of those who record voices for films, adverts, documentaries, etc.” (Orero, 2005, pp. 9-10).

Cabeza Cáceres (2017), who is Spanish, does not specifically mention delivery, but lists voice talent under the heading of sound. Different practices between countries are unlikely to affect interpreting, as the interpreter will not give way to a voice talent, although the context in which the interpreting takes place will affect the importance given to delivery. In 1952, Jean Herbert pointed out that in a literary gathering, elegance of speech might take precedence whereas the nuances of the words would be of utmost importance in interpreting at a diplomatic conference. Equally, accuracy would be more important than euphony in a medical setting. Ultimately, it is important to note that delivery should be a criterion of quality for AD, because oral translations convey as much meaning from the way words are uttered, as from the words themselves.

Lee (2008) goes further in clarifying aspects of delivery for the assessment of performance of consecutive interpreters. First, she suggests ways in which good delivery can be recognized, proposing that the voice needs to attain the “quality of good public speaking” and achieve “successful communication” (Lee, 2008, p. 173). In the rubric of her marksheets, she lists the following examples of what she calls “deviations” from good delivery: “inarticulate speech, pauses, hesitation, false starts, fillers, irritating noises, repetition, excessive repairs or self-correction, unconvincing voice quality and monotonous intonation, and irritatingly slow speech rate.” Next, she defines three bands by which performance can be rated: “Excellent” or “Good delivery with few deviations, and “Poor delivery with some deviations”. This is not only of practical use in an assessment context (see below); it also demonstrates the degree of detail into which broad criteria must be divided if aspects of quality are to be usefully defined and “proven” for the purposes of assessment. However, while the microcriteria contained in Lee’s rubric offer a general lesson for AD, the overlap is not complete. In an AD setting, delivery deviations might also include extraneous noises such as microphone cable noise, script rustles and lip smacks. An excessively slow speech rate may well be irritating, but so might any rate that is inappropriate to the scene. Racing through a lyrical love scene would be as inappropriate as describing a high-octane chase sequence at a snail’s pace. Schjoldager’s (1996) criteria extend Lee’s by asking, “Are there irritating unfinished sentences?”, “Are there irritating outbursts?”, “Is the voice unconvincing?” Again, these are more applicable to AD in some contexts than in others. For example, an unfinished sentence is unlikely to occur in recorded AD, unless the recording has failed due to a technical error; whereas in live situations, it is perfectly possible for a describer to start to describe, only to be cut off by the actor’s dialogue; perhaps because the performance has quickened during its run. An irritating outburst—Schjoldager (1996, Table 1) gives as examples “*ah! oh no! shit! etc.*” — would be equally unprofessional in AD, and both AD and interpreting would be negatively affected if the voice was unconvincing—if, as Schjoldager puts it, the voice “gives the interpreter’s doubts away.”

Language

A description or interpretation that does not use language well will surely fail any test of quality. Lee (2008) places language under the heading of “TL [target language] quality,” explaining that this “refers to the quality of the rendition in terms of linguistic ‘correctness’, naturalness, and contextual appropriateness of language” (p. 169). Again, she lists specific deviations: “incorrect pronunciation, accent, and stress; incorrect grammar; unidiomatic language; interference from the source language; and inappropriate language in the target culture and for the target audience (register misuse).” Her first three items might be better placed under delivery, although Cabeza Cáceres (2017) also places pronunciation under language. Poor pronunciation can break the trust built between the describer and the AD user (Fryer, 2016), as well as potentially confuse the audience. For example, the describer pronouncing the name of a character or a location differently from the way the cast pronounces it can create a lack of coherence. Misplaced stress can do the same. However, it can also be highly useful in AD because of its ability to alter nuances of meaning economically. For example, it can be used to suggest the arrival of a new character or location. “Mr. X takes a seat” would imply a degree of surprise at the presence of Mr. X, leading the audience to assume either that he is a new character or newly arrived in the scene, or perhaps that he is acting out of character or in a way that is not commensurate with his status. Left unstressed: “Mr. X takes a seat” leads us to assume he has been present all the time and sitting is a perfectly natural thing for him to do. Allowing the stress to fall on any one of the words in that sentence might alter users’ understanding of his action.

As noted above, an assessment model based on deviations in language might be seen as unnecessarily negative. In AD, it may be appropriate to positively reward a student for good use of language. Use of language that is vivid and economical is expressly encouraged. For example, in an experimental stylesheet for AD teaching purposes, Remael recommends the use of “adjectives that are specific (‘devious’ vs ‘bad’) and visual or other metaphors (‘a man like a mountain’)” and “vivid verbs that detail the action rather than accumulating verbs + adverbs (‘run’ vs ‘walk quickly’; ‘skip’ vs ‘walk cheerfully’)” (2005, p. 3).

The ability to select the most appropriate words is doubtless why “perfect use of mother tongue” is ranked as the second most important competency in AD in the IO1 questionnaire. This is also a departure from interpreting, where public service interpreters, in particular, need excellent command of two languages (SL and TL).

Grammar

As a subset of poor language use, incorrect grammar, identified by Cabeza Cáceres (2017) as a microcriterion, is also likely to damage the trust a user has in the provider (in either AD or interpreting). However, it might not be appropriate to criticize poor grammar in all AD settings. For example, in AD of dance, practitioners actively recommend abandoning grammatical structures in order to achieve better rhythmic effects (e.g., Fryer, 2009, 2016; Margolies, 2015). It also should be remembered that AD is not written discourse. It does not need to stick to the rules, and should resemble conversation. In particular, it should include the type of contractions common in everyday speech (e.g., “it’s”, “they’re”). Incomplete sentences are common, especially to introduce a location or setting (e.g., “Night”, “The Bar”).

Marzà Ibañez (2010) refers to “syntax and morphology errors,” broad categories that cover a range of grammatical errors. It is worth noting that although AD is an oral form of translation, spelling mistakes

and poor punctuation still matter in an AD script, because, as previously mentioned, a voice talent might deliver the AD. A spelling mistake may lead to a stumble or hesitation, necessitating a re-record. This is expensive in terms of time and studio costs (Fryer, 2016). In live situations, a hesitation or a stumble is even more costly, as it cannot be corrected.

Articles

Cabeza Cáceres (2017) places the use of articles before grammar as microcriteria in his language category. Articles are of particular concern in AD in languages that use them, because the use of a definite article carries the implication that the object or location to which it refers has been encountered before. “She picks up a gun” suggests that a weapon is casually lying around and does not belong to anyone in particular. “She picks up the gun” implies that the gun has already been seen by the sighted audience and will have previously been mentioned in the AD. Misuse of articles is a very common mistake among students of AD. This is perhaps why Cabeza Cáceres lists it first. However, misuse of pronouns can be equally serious. The use of a pronoun where more than one character of a particular gender is present leads to confusion as to who carried out an action. The literature on pronouns is much more developed in interpreting studies than in AD (e.g., Stevenson, Knott, Oberlander, & McDonald, 2000) and describers could usefully learn about constructions that minimise confusion. Cabeza Cáceres places pronouns under Content.

Cabeza Cáceres (2017) includes confusing phrasing as a subsection of grammar. Sometimes it is simply a case of juxtaposition that gives rise to confusion, rather than grammar *per se*. For example, in a scene from *The Pianist* (Polanski & Benmussa, 2002) a student’s AD reads: “Regina looks to the foot of the wall where Jews’ bodies shot down are lined up.” Grammatically it is correct but the juxtaposition of shot down and lined up sounds contradictory and can cause a mental double take.

Marzà Ibañez (2010) covers more grammatical deviations under the broad category of “style,” within which she lists: “Cacophonies, pleonasms, unnecessary repetitions, poor style. Deviates from the style of the text to be described.” Of these, pleonasms are particularly relevant to AD, where economy of expression is encouraged (Remael, 2005). The use of more words than necessary to convey meaning (e.g., see with one’s eyes) is a common, if unfortunate, occurrence in AD, particularly in descriptions of characters where, for example, a woman might be described as having a diamond necklace sparkling around her neck. The inclusion of such redundant information is discouraged for masking more of the soundtrack than necessary during what Braun (2007) refers to as *dynamic* AD. It is also believed to add to the user’s cognitive load in any AD mode, although this has not been tested.

Loyalty

Loyalty is the third of Schjoldager’s macrocriteria (1996). By loyalty she means faithfulness to the ST, which is also the first of four criteria mentioned by Pöchhacker (2001), namely: “accurate rendition”, “adequate target language expression”, “equivalent intended effect” and “successful communicative interaction” (p. 41).

Loyalty in Schjoldager’s sense is elsewhere referred to as accuracy. Lee argues that: “the sense consistency between the original message and interpretations is a fundamental aspect of interpreting quality” (2008, p. 168). However, Lee suggests that accuracy should be regarded as more than a simple correspondence between ST and TT. Like Pöchhacker (2001, p. 41) Lee regards an accurate rendition as one that conveys “an equivalent intended effect” (p. 168). This is returned to below.

Quality Assessment in Audio Description

Like users of interpreters' services (Kurz, 2001; Ng, 1992), most AD users are not in a position to judge the accuracy of the translation. Accuracy in AD includes the possibility not only of misnaming something or using an inappropriate verb (saying someone walks when they run), but also of a false addition (saying something happens when it doesn't) and omission. Omission can be a deliberate strategy both in AD and interpreting, as it is no more possible or desirable to translate every word than it is to describe everything in the visual array. The skill lies in knowing what to leave out. Omission covers a broad spectrum of possibilities—with censorship at one extreme—but again, neither ADLAB PRO nor Cabeza Cáceres (2017) mentions it. Marzà Ibañez (2010, p.149) refers to omission as “suppression”. In addition to: “Suppression of visual, acoustic or textual information needed to understand the text, especially setting the scene (plot, time and space) and the characters in the scene,” she also refers to “Suppression of relevant thematic connections.” This is an area to be treated with caution. The ITC (2000) guidelines suggest that “support information” can help “minimize confusion,” giving the example of clarifying an acronym used in the dialogue. However, the American guidelines caution against including anything in the description that cannot be seen. Returning to the example from *The Pianist*, it would be tempting to insert: “Regina looks to the foot of the wall at a line of bodies of Jews who have been shot while trying to escape.” The logic for not doing so is that if there is no explanation for sighted people, it would be patronizing to include it in the AD for blind people (Clarke, 2007; Fryer, 2016). One of the pleasures of watching a film is teasing out unspoken connections.

Omission is used as an indicator of lack of loyalty in Schjoldager's assessment criteria. She also uses loyalty in a more literal sense in which lack of it would be shown by mocking the speaker or the speaker's message. This is an area that describers joke about among themselves (e.g., she gestures unnecessarily), but the AD literature rarely mentions it directly, although several guidelines warn against expressing an opinion (ITC, 2000; Remael, 2005).

Schjoldager (1996) also assesses the inclusion of any unnecessary changes in the translation.

Equivalence of Impact

There is some agreement amongst interpreting scholars, that quality can be considered in terms of equivalent effect (Déjean le Féal, 1990, p. 155) be that as “a faithful image” (Gile, 1991, p. 198) or “an exact and faithful reproduction” (Jones, 1998, p. 5). One way to judge equivalence of impact is through reception studies. Several (e.g., Fryer & Freeman, 2012a, 2012b; Walczak & Fryer, 2017a, 2017b) have measured impact using the psychological construct of presence (Lombard & Ditton, 1997; Minsky, 1980). Likened to immersion, presence assesses the degree to which an audience member feels transported by the mediated product, such that they are no longer aware of its being mediated (e.g., Biocca & Harms, 2002). Assessing presence usually takes the form of a post hoc questionnaire such as the ITC Sense of Presence inventory (ITC-SOPI) (Lessiter, Freeman, Keogh, & Davidoff, 2001). This requires users to rate statements that measure their levels of engagement; spatial presence (how much they felt they were visiting the places shown on screen); ecological validity (how real it seemed); and negative effects. Fryer and Freeman (2013) argue that one advantage of using presence measures is that the responses of people with and without sight can be directly compared, in contrast to simply asking users about the AD on which the sighted cannot comment and which the users arguably should not notice (Fryer, 2016). Ramos (2015) and Fryer and Freeman (2014) have also turned to other psychological measures to assess emotional impact, using the Emotional Elicitation scale (Gross & Levenson, 1995).

Coherence and Plausibility

The last of Schjoldager's (1996) broad categories is "coherence and plausibility." Braun (2007) and Taylor (2014) make it clear that cohesion or links between utterances are also important if AD is to make sense to the user. In AD, achieving coherence is arguably most challenging in describing visual humour, in that the user needs to understand why the sighted audience is laughing, making not only the discourse coherent, but also the links between the action and the sighted audience's response. Often it is challenging to find either the right words or space within the dialogue to set up the joke. For example, in a production of *Barbershop Chronicles* (Ellams, 2017, dir. B. Sheibani), one character asks the cost of a haircut, and lifts his hat to reveal that he is bald. The audience laughs immediately, leaving post description as the describer's only option (deviating from the criterion of synchrony, see below). Plausibility can also be problematic unless there is time to explain how an effect has been achieved. For example, a character might disappear in the dimming lights, by descending through a trapdoor into the stage, or by being flown out on wires. It is necessary to understand the nature of theatrical illusion for a sudden disappearance to be plausible. Arguably, in the competencies suggested by ADLAB PRO IO1(2017a), this comes under "choosing the most relevant information to describe," which was ranked highest in the IO1 questionnaire. Al-Kharabsheh (2017) argues that relevance is a useful concept for determining quality in interpreting, such that "the degree of quality would rise exponentially with the degree of relevance achieved by the interpreter's TL versión." A similar idea has been proposed for AD based on J. J. Gibson's theory of affordances (Fryer, in press) whereby the elements chosen by the describer reflect what the ST "affords" the sighted audience. For example, a horror film affords the opportunity to be scared in a safe setting, while a farce affords us the chance to laugh. As stated above, this is a way of ensuring that an AV (audiovisual) product achieves its *skopos* or purpose. Drawing on the field of data quality, Greco (2017) provides a global definition of quality in AVT as providing information that is "fit for purpose."

Synchrony

Both Marzà Ibañez (2010) and Cabeza Cáceres (2017) include timing as an indicator of AD quality. AD, like other AVT modes such as dubbing and subtitling, necessitates the production of a timed text. Marzà Ibañez subdivides timing into "synchrony," (AD should be synchronized with action and should not break the suspense by giving away information too late or too early) and "pace" (the AD script is criticized if it is too long for the "gap" identified in the dialogue, or is too concentrated). Although pace has been mentioned under delivery above, the decision of where to position an utterance in the soundtrack is independent of how it is delivered. Density should also be treated as a separate matter. A script is of poor quality not only if too much, but also if too little has been written for the specified duration, as this results in the soundtrack being masked unnecessarily. The solution would be to shorten or lengthen the specified gap (within the constraints of the soundtrack), rather than to speak more quickly or slowly. Synchrony has also been recognised as a constituent element in interpreting. Pöchhacker (2004, p. 117) defines synchrony as the "décalage" or time lag between the original speech and the interpreter's rendition, recognising that better coherence is achieved as lag decreases. Alonso Bacigalupe (1999) stresses that in AV presentations, the increasing need to translate visually presented information such as PowerPoint increases the interpreting lag. Baxter (2016) recognises that with the proliferation of English as a lingua franca, there is greater pressure on the interpreter to remain in synchrony with the ST, as the audience is more likely to notice any discrepancy. Similarly, although AD is written keeping in mind people

Quality Assessment in Audio Description

who are totally blind, it must be recognised that 93% of the visually impaired population (RNIB, n.d.) retains some residual vision and likewise needs any lag to be minimal. ADLAB PRO places timing as one of the technical competencies in AD training, ranked fifth out of nine. However, the 100 AD users who completed the survey, ranked “The audio description is not well synchronized with the dialogue and sound effects or with the images” as their chief cause of dissatisfaction with existing AD (2017b).

Divergence in Competencies

AD is what Benecke (2014) calls “partial translation,” with users receiving important source information from the soundtrack, around which the AD is woven, whereas conference interpreters are free to speak over the ST with little fear of reducing their users’ understanding aside from the proliferation of English mentioned above. For interpreters, the significance of timing will be dependent on the mode of interpreting employed (simultaneous or consecutive). It is worth noting that the professional bodies of those two types of translation activities have some differing quality criteria, but share content (structure/delivery (fluency) and language expression (vocabulary/grammar/style).

The IO1 Questionnaire also demonstrates the importance that experienced describers and teachers of AD attach to understanding the target culture, as “knowledge of the needs of the visually impaired” was ranked third. It is interesting that none of the sources for the interpreting criteria discussed so far make reference to acquisition of such cultural knowledge. However, Gile (2001) cites research by Hearn et al. (1981), which recognizes that knowledge of both source and target language, and “of the migrant culture,” to be among the qualities that makes a good interpreter. The assessment criteria drawn up for the European Masters in Conference Interpreting (EMCI) include “cultural comprehension” under content.

Multimodality

Mari Carmen Campoy-Cubillo (2017) recognizes that traditional assessment criteria for interpreting concentrate almost exclusively on audio content, overlooking the multimodal nature of most speech acts, where nonverbal acts such as facial expression, eye gaze, and gesture are also important, in addition to prosody (discussed above.) It is ironic that experienced audio describers are extremely alert to such multimodal elements, which constitute a large proportion of their descriptive utterances, since non-verbal forms of communication, being inaudible, are inaccessible to AD users. Yet, description itself is resolutely unimodal, as everything is turned into audio. There are exceptions to this in some AD modes. For example, touch tours are common in live theatre AD; workshops often precede a dance AD; and touch objects or tactile images may be included in a described tour of a museum. However, any attempt to include multimodality as an assessment criterion for AD in general would not be appropriate. Wadensjö (1998) acknowledges that quality criteria are context dependent for interpreting: “In practice, there are no absolute and unambiguous criteria for defining a mode of interpreting which would be ‘god’ across the board” (p. 287). The same is surely true for AD.

Stakeholders in Quality

In a similar vein, another area in which research in interpreting has led the way is in highlighting the many stakeholders who have varying views on what constitutes quality. Stakeholders in AD quality include:

- **AD users** who need to know they're getting an enjoyable and reliable description
- **The service provider (SP)** who needs to know that they're providing a good service (whether or not to use those describers again)
- **The SP's customers or funders** (e.g. Arts Council England) especially where accessible provision is a funding requirement
- **The Law** (is the service provider meeting its obligations under the Equality Act or equivalent?)
- **The description agency** –who employs the describers– needs to know that they're providing a good service (whether or not to use those describers again)
- **The AD student** (and the professional) who need to know if they're doing a good job.
- **The ST creator(s)** (including the director, script writer or devisors) who need reassurance that their product is not distorted in the accessible version.
- **ADLAB PRO** (or any course provider) who needs to know if their students are reaching industry standards.

Pöchhacker (2001) points out that: “There is agreement in the literature across the typological spectrum that interpreting, conceived of as the task of mediating communication between interactants of different linguistic and cultural background, is, first and foremost, a service designed to fulfil a need” (p. 421). As those needs vary, so too will the criteria for determining quality. Lee (2008) highlights the difference between academic and vocational courses: “given that the focus in performance assessment of interpreting students lies in the product quality rather than the service quality” (p.168). As argued above, a distinction must be drawn between ways of assessing the quality of the description, and assessing the skills and competence of the describer. For interpreting, Pöchhacker (2001) defines this as the difference between “product-orientation” and “interaction-orientation” (p. 412), while Kurz (2001) distinguishes between quality from the perspective of the speaker, and quality from the perspective of the listener or user. It is easy to see that what blind or partially sighted users think of as quality might differ from the view of the actors, the describers, or the director of the product being described. Approaches such as auteur AD (Szarkowska, 2013), integrated AD (Cavallo, 2016) or accessible filmmaking (Romero-Fresco, 2013) seek to merge at least some of these views, to improve the quality by creating AD from a multidimensional perspective.

Quality From a Marketing Perspective

Applying a marketing model in relation to quality in interpreting, Kurz (2001) suggests: “Quality must begin with customer needs and end with customer perception” (p. 394). In that case, it is important to define the term “customer.” According to the Business Dictionary (n.d.) a customer is “a party that receives or consumes products (goods or services) and has the ability to choose between different products and suppliers.” Although the AD user might seem to be the customer, at no stage does an AD user attending a live performance have choice over which AD provider to use. That is a decision already made by the venue, the producer or the company, whichever has taken responsibility for making the performance accessible. Where contracts exist, they are made between the venue and the individual describer, or perhaps between the theatre and the agency responsible for providing the AD. The use of agencies (or service providers) separates the AD user and the describer still farther. For example, according to their website, the UK charity VocalEyes (VE) (vocaleyeyes n.d.) provides “opportunities for blind and partially sighted people to experience the arts at UK’s theatres, museums, galleries and heritage.” VE would be contracted

Quality Assessment in Audio Description

by a theatre or a museum to provide a description, and in turn would appoint one or two describers to carry out the task. The direct relation between customer and service provider becomes indirect. In the case of live AD, the venue is the customer of the agency that, in turn, is the customer of the describer (or at least pays the describer for their services). Under the marketing model, it would be the theatre's responsibility to determine the quality of the AD. It would also be advisable (and is, in fact, the case) for the agency to have its own quality control (QC) mechanism governing the describers it engages. To use a concrete example, VE has several stages of quality control embedded in its working practices.

Embedded Quality Control

No matter how experienced they may be, VE describers rarely work alone. This is to ensure that a consistent standard is maintained across VE descriptions. In most (but not all) cases, two describers will be allocated to describe a production, so that any deficiency or subjectivity in one describer will be compensated for and corrected by the other. At the first stage they will work together to produce an audio introduction (AI) that is recorded and distributed in advance of the described performance. An AI is defined as “a piece of continuous prose, spoken by a single voice or a combination of voices and lasting between 5 and 15 minutes [that] provides relevant information from the printed programme, including running time, cast and production credits, as well as detailed descriptions of the set, costumes, and characters” (Romero-Fresco & Fryer 2013, p. 289). The provision of an AI could be one benchmark of a good quality description. Its purpose is to provide context and more detail than time allows during the *dynamic* AD. This is clearly another aspect in which AD parts company from interpreting, the act of which is restricted to the live element. The second stage of QC in the preparation of the AI comes as the draft AI is forwarded to another describer (the VE editor), who has not seen the production. This means the editor assumes the role of an AD user, to the extent that her understanding of the content is not influenced by sight. The VE editor will review the AI before it is recorded, correcting it where necessary in consultation with its authors. At the recording stage, the AI can be further modified, if it is deemed necessary, by the recording producer, who is himself an AD user, or by one of the two describers who are voicing the recording (this is occasionally, although rarely, the same describer(s) that wrote the text). The aspects of quality that will be formally and informally assessed at the stages outlined so far will be accuracy and omission (monitored by the two describers who initially saw the show and wrote the AI); clarity, comprehensibility, grammatical accuracy, and *euphony* or the quality of being pleasing to the ear, (assessed by the editor, the producer, and those describers who are voicing the recording). As for the *dynamic* AD, QC is embedded in the “dry run,” which is a live testing of the description at a performance shortly before the described performance itself. The AD at the dry run will be listened to only by the describers themselves, who are responsible for assessing each other's AD, and perhaps by a theatre technician who will assess the sound quality from a technical perspective. Occasionally a third describer will also attend to take notes and give feedback. These notes will generally relate to accuracy, omission, ambiguity (for example, using a pronoun inappropriately), and to the technical quality of the delivery system. In some cases, (for example, at the UK's National Theatre), feedback is also solicited from a visually impaired person attending the dry run, who will give notes on any aspect of the description that made the production confusing or inaccessible to them.

The final internal quality control (QC) system embedded in a VE AD takes the form of a yearly formal assessment of every describer. A fellow describer and a visually impaired user will attend an AD

performance (including the touch tour that precedes the performance). Afterward they will complete a semistructured pro forma that includes the following questions:

- **Script Content:** Did the describer identify the characters well and consistently? Were there any moments where you felt the description became judgmental or opinionated? Was the description accurate? Was it easy to understand? Did the describer translate visual images into words that were descriptive and in keeping with the narrative?
- **Delivery:** Did the describer time the description so as not to interrupt the flow of dialogue? Did he/she make good use of music and sound effects? Was the delivery sensitive to the scene? Could you hear everything clearly? Was the delivery well paced? Did you feel comfortable that the description would give you all the information you would need? Were there any issues of pronunciation? Did you hear any inappropriate sounds such as the rustle of a script, microphone noises etc.?

There is also a heading marked “Other” to allow for feedback that did not fall under any specific category. The questions for the visually impaired assessor also ask about the quality of the recorded AI that was received before attending the performance. For the *dynamic* AD, the category headings are: Vocal Quality, Scene Sympathy, Timing, Content, Interpretation, and Confidence. The final section is titled Front of House, and deals with practical aspects of the theatre visit, such as booking tickets, arriving at the theatre, picking up headsets and tickets, and treatment by the theatre staff. It is worth noting that the AD user is paid to provide this assessment, and undertakes training before doing so. This is in recognition of the fact that an informed, formal assessment process is different from general user feedback. Also, a person who is attending a performance for his or her own enjoyment does not necessarily want to provide a detailed response at the end. Nevertheless, informal feedback is also helpful. As Marrone (1993) notes for interpreting, “[c]hecking our own assumptions against our listeners’ feedback may provide useful orientation for practitioners, teachers and aspirant interpreters” (p. 35). The same is surely true in the field of AD. Informal feedback may be provided to the theatre by the front of house staff, with whom the AD users are most likely to communicate as they hand back their headset or leave the venue at the end of a performance. It is up to the theatre whether or not such feedback is circulated to the agency or to the describers. However, it doubtless influences the decision of the theatre as to whether to use the same agency or describers again.

As with people using interpreters (Kurz, 2001), one difficulty is that peripheral aspects of the customer experience may sway AD users’ opinions. In theatre AD, the ease of getting to the venue, the friendliness of the staff, guide dog facilities, the efficacy of the technology used to receive the description, and the source content (the play itself) may all have an impact on user experience, and hence affect judgement of the quality of the AD. This is why such aspects have been separated out in the VE assessment form. It has also been argued elsewhere (Fryer, 2016) that the best AD is that which is barely noticed.

Weighting

As the previous paragraphs make clear, there are many contributing aspects to producing a quality AD. The difficulty comes not only in isolating those ingredients, but also in weighting the importance of the contribution each makes to the final product.

Questionnaires such as the IO1 for AD and the surveys carried out for interpreting by Bühler (1986), Kurz (2001), Pöchhacker (2000), and Garber and Maufette-Leenders (1997), among others, offer meth-

Quality Assessment in Audio Description

odological solutions whereby many respondents rank criteria, triangulating multiple points of view. This introduces a degree of objectivity into what is otherwise a subjective decision, while making clear that there is no single view of what constitutes quality nor what it takes to produce a quality product.

To achieve something similar for AD, ADLAB PRO (2017b) has prepared a second questionnaire (IO2) that seeks to replicate interpreting surveys, by asking users, practitioners, and providers to rate 10 aspects of AD quality, as well as giving them the opportunity to mention other aspects that might be missing from the list. It is expected this will highlight a similar divergence of stakeholder views. The specific aspects to be rated are shown below:

To deliver a good audio description, an audio describer should be able to:

- Express meaning succinctly
- Select significant information
- Use language that sparks the imagination
- Compile an audio introduction
- Provide the listener with a way of “seeing” what is described
- Provide the listener with a way of “understanding” what is described
- Possess an excellent command of mother tongue
- Use non-ambiguous language
- Use language that is suited to the product
- Use language that is suited to the audience
- Other please specify

SOLUTIONS AND RECOMMENDATIONS

In contrast to the mark scheme proposed by Cabeza Cáceres (2017), Lee (2008) provides a useful alternative, offering grading rubrics or qualitative statements against which a student’s holistic performance can be assessed (p. 172). First, she gives a broad definition of three main areas (TL quality, delivery, and accuracy). and then she lists examples of deviations. Finally, she offers six bands, each with a qualitative illustration of the learner’s achievement ranging from band 1 where it was barely achieved at all to band 6—where it is fully achieved. The same criteria that are used to assess the quality of the translation (e.g. accuracy) are used to reflect on the ability of the translator (e.g. listening ability) again from band 1 —*Very limited understanding of the message of the original speech* to band 6 —*Complete understanding of the message of the original speech*. Marks are awarded according to the band, for each of three category headings: Accuracy (6), Language (6), and Delivery (3), such that the learner can be given a final score out of 15, and the higher the score, the better the quality. This system weights the different elements (delivery attracting fewer possible marks than accuracy or language), while providing useful feedback for pedagogical purposes. Lee’s scoring system would alleviate the problems outlined above of increased (negative) scoring associated with an increase in the duration of the translated product. A space has been created on the form to allow the assessor to cite a reference for where a particular deviation occurred. This would provide the type of evidence required for the describer to learn from their mistakes.

It is worth noting that for various scholars, the microcriteria are placed under different broad headings or macrocriteria. This is of purely semantic concern. It is the microcriteria that enable a judgement to be

made about the skills and competencies of the learner and the extent to which they contribute towards the quality of the final product.

In the appendix to this chapter, Lee's marksheets have been adapted for use in AD, with the addition of the microcriteria and areas unique to AD identified above. Four macrocriteria have been selected: Accuracy (loyalty and impact), Language, Delivery, and Synchrony. They have been left unweighted, meaning the total score would be calculated out of 24. Content was abandoned as a macrocriterion on the grounds that all the aspects could be deemed to constitute content, likewise Lee's category of TL quality.

For Interpreting, Alonso Bacigalupe (2013) proposes a formula of quality that lies in "the interaction between: a) needs and expectations; b) constraints, and c) skills" (p. 18) pertaining respectively to 1) the user; 2) the assignment, and 3) the translator.

Ultimately, all the microcriteria discussed above contribute to the quality of the AD product according to a global criterion, namely the extent to which it fulfils its function to make accessible to its users everything that is available to a sighted audience, while retaining its impact.

FUTURE RESEARCH DIRECTIONS

The partners in ADLAB PRO will test the marksheets included in the appendix. The mass of data collected for IO2 (ADLAB PRO, 2017b) are due to be further refined. These developments might suggest whether weighting should be introduced. Reception studies testing the impact of cognitive load might also introduce new criteria concerning density and language use. A more succinct version of the assessment model needs to be developed and tested for professional descriptions to be used by agencies, service providers or other customers of AD services.

CONCLUSION

It is evident that when it comes to quality, the field of AD, both as a profession and as an academic discipline, faces many of the same challenges as interpreting. Given so much common ground, it seems sensible for AD scholars interested in quality and assessment to make use of existing quality research in the field of interpreting studies. Having said that, it is also important to recognize areas where the fit is incomplete, such as the need for the AD to be considered not in isolation, but in relation to the existing audio elements of the ST. This chapter is a first step toward doing so.

ACKNOWLEDGMENT

This research was supported by the Erasmus+ Programme of the European Union: Project name ADLAB PRO; Project number: 2016-1-IT02-KA203-024311.

The information and views set out in this chapter are those of the author and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

REFERENCES

- ADLAB PRO. (2017a). *Report on IO1: Assessment of current AD training practices*. Retrieved from <http://www.adlabproject.eu>
- ADLAB PRO. (2017b). *Report on IO2: Audio description professional: Profile definition*. Retrieved from <http://www.adlabproject.eu>
- ADLAB PRO. (n.d.). Retrieved from <https://adlabpro.wordpress.com>
- Alonso Bacigalupe, L. (2013). Interpretation quality: From cognitive constraints to market limitations. In R. Barranco-Droege, E. Macarena Pradas Macías, & O. García Becerra (Eds.), *Quality in interpreting: Widening the scope* (Vol. 2, pp. 9–33). Granada: Comares.
- Audio Description Association. (2004). *Certificate in Audio Description Skills Training Course Manual* (unpublished).
- Baxter, R. N. (2016). Aproximación a los posibles efectos del uso de presentaciones visuales sobre la sincronía y el décalage en la interpretación simultánea. *Sendebarr*, 27, 9–23.
- Benecke, B. (2014). *Audiodeskription als partielle Translation. Modell und Methode*. Berlin: LitVerlag.
- Biocca, F., & Harms, C. (2002). Defining and measuring social presence: Contribution to the networked minds theory and measure. *Proceedings of PRESENCE*, 2002, 7–36.
- Braun, S. (2007). Audio description from a discourse perspective: A socially relevant framework for research and training. *Linguistica Antverpiensia, New Series. Themes in Translation Studies*, 6, 357–369.
- Cabeza Cáceres, C. (2017, March). *Professional audio descriptions: Same film, same assignment = same quality?* Paper presented to Advanced research seminar in audio description (ARSAD), Barcelona, Spain. Retrieved from <http://grupsderecerca.uab.cat/arsad/sites/grupsderecerca.uab.cat/arsad/files/cabeza-cacerescristobal-arsad2017.pdf>
- Campoy-Cubillo, M. C. (2017). Listening skills: Issues in assessment and implementation. In V. Bon-signori & B. C. Camiciottoli (Eds.), *Multimodality across communicative settings, discourse domains and genres* (pp. 14–37). Newcastle-upon-Tyne, UK: Cambridge Scholars Publishing.
- Cavallo, A. (2016). Audio description: Ramps on the moon. *Exeunt*. Retrieved from <http://exeuntmagazine.com/features/audio-description-ramps-moon/>
- Clark, J. (2007). *Comments on U.K. guidelines on audio description*. Retrieved from <http://joelclark.org/access/description/ukad.html>
- Déjean Le Féal, K. (1990). Some thoughts on the evaluation of simultaneous interpretation. In D. Bowen & M. Bowen (Eds.), *Interpreting. Yesterday, today, and tomorrow* (pp. 154–160). Binghamton, NY: SUNY. doi:10.1075/ata.iv.27lef
- Fryer, L. (2016). *An introduction to audio description: A practical guide*. London: Routledge.

- Fryer, L. (in press). Perception and psycholinguistics in AVT. In L. Perez-Gonzalez (Ed.), *The Routledge handbook of audio visual translation*. London: Routledge.
- Fryer, L., & Freeman, J. (2012a). Cinematic language and the description of film: Keeping AD users in the frame. *Perspectives: Studies in Translatology*. doi:10.1080/0907676X.2012.693108
- Fryer, L., & Freeman, J. (2012b). Presence in those with and without sight: Implications for virtual reality and audio description. *Journal of Cyber Therapy and Rehabilitation*, 5, 15–23.
- Fryer, L., & Freeman, J. (2013). Visual impairment and presence: Measuring the effect of audio description. In H. Witchel (Ed.), *Inputs-Outputs Conference. Proceedings: Interdisciplinary approaches to causality in engagement, immersion and presence in performance and human-computer interaction* (pp. 1-5). Brighton, UK: Academic Press.
- Fryer, L., & Freeman, J. (2014). Can you feel what I'm saying? The impact of verbal information on emotion elicitation and presence in people with a visual impairment. In A. Felhofer & O. D. Kothgassner (Eds.), *Challenging presence: Proceedings of the international society for presence research, 15th International Conference on Presence* (pp. 99 – 107). Vienna: Facultas.
- Fryer, L. E. (2009). *Talking dance: The audio describer's guide to dance in theatre*. London: Audio Description Association.
- Garber, N., & Mauffette-Leenders, L. A. (1997). Obtaining feedback from non-English speakers. In S. E. Carr, R. Roberts, A. Dufour, & D. Steyn (Eds.), *The critical link: Interpreters in the community* (pp. 131–146). Amsterdam: John Benjamins. doi:10.1075/btl.19.15gar
- Gile, D. (1991). A Communication-Oriented Analysis of Quality. In M. L. Larson (Ed.), *Translation: Theory and practice* (pp. 188-200). Binghamton, NY: SUNY.
- Gile, D. (2001). L'évaluation de la qualité de l'interprétation en cours de formation. *Meta: Journal des traducteurs/ Meta: Translators' Journal*, 46(2), 379–393.
- Gouadec, D. (2010). Quality in translation. In Y. Gambier & L. Van Doorslaer (Eds.), *Handbook of translation studies* (Vol. 1, pp. 270–275). Amsterdam: John Benjamins. doi:10.1075/hts.1.qual
- Greco, G. M. (2017). *Understanding media accessibility quality (UMAQ)*. Paper presented at Media Accessibility Platform International Conference (MAPIC), Vigo.
- Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition and Emotion*, 9(1), 87–108. doi:10.1080/02699939508408966
- Herbert, J. (1952). *The Interpreter's Handbook*. Geneva: Georg.
- Iglesias Fernández, E., Martínez Martínez, S., & Chica Núñez, A. J. C. (2015). Cross-fertilization between reception studies in audio description and interpreting quality assessment: The role of the describer's voice. In R. Baños Piñero & J. Díaz Cintas (Eds.), *Audiovisual translation in a global context. Mapping an ever-changing landscape* (pp. 72–95). Palgrave Macmillan, UK.
- ITC. (2000). *ITC guidance on standards for audio description*. Retrieved from http://www.ofcom.org.uk/static/archive/itc/itc_publications/codes_guidance/audio_description/index.asp.html

Quality Assessment in Audio Description

- Jones, R. (1998). *Conference interpreting explained*. Manchester, UK: St. Jerome Publishing.
- Kalina, S. (2005). Quality assurance for interpreting processes. *Meta: Journal des traducteurs/ Meta: Translators' Journal*, 50(2), 768–784.
- Kohn, K., & Kalina, S. (1996). The strategic dimension of interpreting. *Meta: Journal des traducteurs/ Meta: Translators' Journal*, 41(1), 118–138.
- Kurz, I. (1989). Conference interpreting: User expectations. In D. L. Hammond (Ed.), *Coming of age*. Proceedings of the 30th annual conference of the American translators association (pp. 143-148). Medford, NJ: Learned Information.
- Kurz, I. (1993). Conference interpretation: Expectations of different user groups. *The Interpreter's Newsletter*, 5, 13–21.
- Kurz, I. (2001). Conference interpreting: Quality in the ears of the user. *Meta: Journal des traducteurs Meta:/Translators' Journal*, 46(2), 394–409.
- Lee, J. (2008). Rating scales for interpreting performance assessment. *The Interpreter and Translator Trainer*, 2(2), 165–184. doi:10.1080/1750399X.2008.10798772
- Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence (Cambridge, Mass.)*, 10(3), 282–297. doi:10.1162/105474601300343612
- Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2), 1–39.
- Margolies, E. (2015). Going to hear a dance: On audio describing. *Performance Research*, 2(6), 17–23. doi:10.1080/13528165.2015.1111044
- Marrone, S. (1993). Quality: A shared objective. *The Interpreters' Newsletter*, 5, 35–41.
- Marzà Ibañez, A. (2010). Evaluation criteria and film narrative. A frame to teaching relevance in audio description. *Perspectives: Studies in Translatology*, 18(3), 143–153. doi:10.1080/0907676X.2010.485682
- Matamala, A., & Remael, A. (2015). Audio-description reloaded: An analysis of visual scenes in 2012 and Hero. *Translation Studies*, 8(1), 63–81. doi:10.1080/14781700.2014.943678
- Minsky, M. (1980). Telepresence. *Omni (New York, N.Y.)*, 2, 45–51.
- Ng, B. C. (1992). End users' subjective reaction to the performance of student interpreters. *The Interpreters' Newsletter*, 1(Special Issue), 35–41.
- Orero, P. (2005). Audio description: Professional recognition, practice and standards in Spain. *Translation Watch Quarterly*, 1(Dec), 7–18.
- Pöhhacker, F. (1994). Quality assurance in simultaneous interpreting. In C. Dollerup, & A. Lindegaard (Eds.), *Teaching translation and interpreting 2: Insights, aims and visions. Papers from the Second Langauge International Conference, Elsinore* (pp. 233-242). Amsterdam: John Benjamins.

- Pöchhacker, F. (2001). Quality assessment in conference and community interpreting. *Meta: Journal des traducteurs / Meta: Translators' Journal*, 46(2), 410–425.
- Pöchhacker, F. (2004). I in TS: On partnership in translation studies. In C. Schäffner (Ed.), *Translation research and interpreting research: traditions, gaps and synergies* (pp. 104–115). Clevedon, UK: Multilingual Matters.
- Pöchhacker, F. (2016). *Introducing interpreting studies*. London: Routledge.
- Polanski, R., & Benmussa, R. (Producers) & Polanski, R. (Director). (2002) *The Pianist* [Motion Picture]. UK, Focus Features.
- Ramos, M. (2015). The emotional experience of films: Does Audio Description make a difference? *The Translator*, 21(1), 68–94. doi:10.1080/13556509.2014.994853
- Remael, A. (2005). *Audio description for recorded TV, cinema and DVD. An experimental stylesheet for teaching purposes*. Retrieved from <http://www.hivt.be>
- RNIB. (n.d.). Retrieved from <https://www.rnib.org.uk/rnibconnect/how-i-see>
- Romero-Fresco, P. (2013). Accessible filmmaking: Joining the dots between audiovisual translation, accessibility and filmmaking. *The Journal of Specialised Translation*, 20, 201–223.
- Sacks, O. W. (1998). *The man who mistook his wife for a hat and other clinical tales*. London: Gerald Duckworth.
- Schjoldager, A. (1996). Assessment of simultaneous interpreting. In C. Dollerup & V. Appel (Eds.), *Teaching translation and interpreting 3: New horizons* (pp. 187–195). Amsterdam: John Benjamins. doi:10.1075/btl.16.26sch
- Shebani, B. (Director). (2017). *Barbershop Chronicles*. National Theatre of Great Britain.
- Stevenson, R., Knott, A., Oberlander, J., & McDonald, S. (2000). Interpreting pronouns and connectives: Interactions among focusing, thematic roles and coherence relations. *Language and Cognitive Processes*, 15(3), 225–262. doi:10.1080/016909600386048
- Szarkowska, A. (2013). Auteur description: From the director's creative vision to audio description. *Journal of Visual Impairment & Blindness*, 107(5), 383.
- Taylor, C. (2014). Intertextuality. In A. Maszerowska, A. Matamala, & P. Orero (Eds.), *Audio description: New perspectives illustrated* (Vol. 112, pp. 29–40). Amsterdam: John Benjamins.
- Torres-Simón, E., & Pym, A. (2016). The professional backgrounds of translation scholars. Report on a survey. *Target. International Journal of Translation Studies*, 28(1), 110–113. doi:10.1075/target.28.1.05tor
- Vercauteren, G. (2007). Towards a European guideline for audio description. In J. Díaz Cintas, P. Orero, & A. Remael (Eds.), *Media for all: subtitling for the deaf, audio description and sign language* (pp. 139–150). Amsterdam: Rodopi.

Quality Assessment in Audio Description

Vocaleyes. (n.d.). Retrieved from <http://vocaleyes.co.uk>

Wadensjö, C. (1998). *Interpreting as interaction*. London: Longman.

Walczak, A., & Fryer, L. (2017a). Creative description: The impact of audio description style on presence in visually impaired audiences. *British Journal of Visual Impairment*, 35(1), 6–17. doi:10.1177/0264619616661603

Walczak, A., & Fryer, L. (2017b). Vocal delivery of audio description by genre: measuring users' presence. *Perspectives*, 1-15.

Wissmath, B., Weibel, D., & Groner, R. (2009). Dubbing or subtitling? Effects on spatial presence, transportation, flow, and enjoyment. *Journal of Media Psychology*, 21(3), 114–125. doi:10.1027/1864-1105.21.3.114

KEY TERMS AND DEFINITIONS

Audiodescription (AD): The practice of translating visual information into verbal information to make the arts accessible to people who are blind or partially sighted.

Blind: An adjective denoting the possession of a sensory disability (usually acquired) that severely limits the reception of visual information in respect either of visual field or acuity of sight. Blindness is a spectrum that, at its most severe, means being unable to distinguish light and shade.

Delivery: The reading aloud of the AD script during recording or performance.

Partially Sighted: An adjective denoting the possession of a sensory disability (usually acquired) that limits the reception of visual information to some degree, in respect to either visual field or acuity of sight.

Pro Forma: A standard form or document.

VocalEyes: A UK AD charity funded by Arts Council England, whose mission is to increase the opportunities for blind and partially sighted people to experience and enjoy art and heritage.

APPENDIX

AD Assessment Rating Scales (After Lee 2008)

Table 4.

Criterion	Band	Effectiveness of Visual Information Conveyed	AD Ref.	Ability
Accuracy The quality of faithfully conveying the visual information of the ST with semantic and pragmatic equivalence i.e. reproducing the same meaning and intended effect. Deviations from the ST should be considered in terms of the effect on the coherence/logic and faithful rendering of the message Examples of deviations: omissions, additions, and unjustifiable changes of the meaning; failure to convey the emotion or visual impact of the scene; poor use of pronouns leading to confusion.	6	The visual information was delivered accurately with intended effect.		Complete understanding of the needs of users
	5	The visual information was generally delivered with intended effect but a few minor deviations from the source text were found, which did not significantly affect the overall meaning or coherence		Good understanding of the needs of users
	4	The visual information was mostly delivered. Some deviations from the source text with an impact on the meaning and effect but coherence was maintained.		Adequate understanding of the needs of users
	3	The message was delivered inaccurately with many deviations from the source text and coherence was compromised.		Inadequate understanding of the needs of users
	2	The message was delivered inaccurately with serious deviations from the source text and incoherence.		Poor understanding of the needs of users
	1	The interpreted message was incoherent and completely inconsistent with the source text.		Very limited understanding of the needs of users
	Mark		/6	

Table 5.

Criterion	Band	Linguistic Quality of AD	AD Ref	Ability
Language The quality of faithfully conveying the message of the ST with semantic and pragmatic equivalence i.e. reproducing the same meaning and intended effect. Deviations from the ST should be considered in terms of the effect on the coherence/logic and faithful rendering of the message Examples of deviations: omissions, additions, and unjustifiable changes of the meaning; uneconomic use of language (pleonasm); clumsy language use (cacophony); grammatical errors; inappropriate use of pronouns	6	Excellent use of language with few linguistic errors and vivid & appropriate expressions.		Excellent language proficiency
	5	Very good use of language with a few minor linguistic errors that do not hinder immediate appreciation of the ST, using appropriate expressions.		Very good language proficiency
	4	Good use of language with a few linguistic errors that may hinder immediate comprehension, but quite understandable. A few minor inappropriate target language expressions were found.		Good language proficiency
	3	Adequate use of language with some linguistic errors that hinder comprehension and some inappropriate target language expressions		Adequate language proficiency
	2	Inadequate use of language with many linguistic errors and inappropriate target language expressions were consistently found.		Inadequate language proficiency
	1	Poor target language production with inappropriate target language expressions.		Poor language proficiency
	Mark		/6	

Quality Assessment in Audio Description

Table 6.

Criterion	Band	Quality of AD Delivery	AD Ref.	Ability
Delivery Quality of good public speaking; successful communication; effective improvisation if necessary. Examples of deviations: inarticulate speech, pauses, hesitation, false starts, fillers, irritating noise, repetition, excessive repairs or self-correction, unconvincing voice quality and monotonous intonation, lipsmacks or other mouth noises, mic. cable noise & irritatingly slow/fast speech rate.	6	Excellent delivery with few deviations		Excellent presentation/communication
	5	Very good delivery with very few deviations		Very good presentation/communication
	4	Good delivery with a few deviations		Good presentation/communication
	3	Adequate delivery		Adequate presentation/communication
	2	Inadequate use of language with many linguistic errors and inappropriate target language expressions were consistently found.		Inadequate presentation/communication
	1	Poor delivery with some deviations		Poor presentation/communication
	Mark		/6	
Criterion	Band	Quality of AD Timing		Ability
Synchrony The AD was successfully woven around the soundtrack Examples of deviations: talking over dialogue or important sound effects; describing too soon or too late (asynchronously); giving away the plot	6	Excellent timing with few deviations		Excellent timing
	5	Very good timing with few deviations		Very good timing
	4	Good timing with a few deviations		Good timing
	3	Adequate timing with a few deviations		Adequate timing
	2	Inadequate timing with frequent deviations.		Inadequate timing
	1	Poor timing with frequent deviations.		Poor timing
	Mark		/6	
Total Mark			/24	

Section 3

Process–Oriented Translation Quality Assessment

Chapter 8

Process–Oriented Assessment of Problems and Errors in Translation: Expanding Horizons Through Screen Recording

Erik Angelone
Kent State University, USA

ABSTRACT

Screen recording has gradually emerged as an efficacious tool in the context of process-oriented translator training. From an assessment standpoint, process protocols derived from screen recordings would seem to hold great potential as an empirical means through which translators and translation trainers can re-trace errors found in translation products back to underlying problem triggers that emerge during their creation. This chapter will begin by outlining how screen recordings can be utilized to reverse engineer translation products for purposes of process-oriented assessment. A series of directly observable indicators will be linked with various error classification parameters, including locus (comprehension, transfer, or production), phase (drafting or revision), and information retrieval type (internal or external) in providing assessors with a diagnostic gauge for pinpointing potential triggers. The chapter will conclude with some preliminary data on evidence of inter-rater consistency when screen recording is used in such a diagnostic capacity by various student populations.

INTRODUCTION

In the context of translation pedagogy, from the mid-1990s onwards, learner self-assessment has gained momentum as trainers have come to embrace the constructivist tenets of guiding students to “assume responsibility for their own work” and “assess the quality of their own performance” (Kiraly, 2003, p. 31). This trend is driven by a number of interrelated factors. From the perspective of language industry realities, where formal external feedback tends to be scant and no news is usually good news, trainers

DOI: 10.4018/978-1-5225-5225-3.ch008

realize that providing students with frameworks and a skillset to assess their own performance prior to entering the industry becomes crucial. If professional translators have aspirations of advancing along an expertise trajectory, deliberate practice is an absolute necessity, and one of the primary conditions that needs to be met in order for practice to be deliberate involves working with feedback (Ericsson, 1996, p. 21). While ISO 17100 (ISO, 2015) establishes an ideal best practice of translations passing through stages of editing, proofreading, and other revisions for purposes of quality assurance, conditions are not in place for the translators to necessarily receive concrete, informative feedback on their performance. ISO 17100 also calls for the translator to check his/her own work after finalizing the initial translation. In many situations, particularly in those involving translation work that does not meet ISO 17100 specifications, the onus for quality assurance is predominantly on the translator. Oftentimes, professional translators need to be in a position to assess their own performance, and their being able to successfully do so hinges upon their receiving formal training in this area. The importance of this is reflected in translation competence models foregrounding self-evaluation as an assessment instrument (Hurtado Albir, 2015, p. 269).

A second factor contributing to the current trend towards learner self-assessment involves technological advancement. Elearning management systems have opened new doors insofar as opportunities for learner collaboration and peer-based assessment are concerned, particularly in combination with tutor moderation (Robinson, López Rodríguez, & Tercedor Sanchez, 2006). Blogs, wiki projects, and rubric templates have expanded the horizons regarding not only from whom students of translation receive feedback, but also in which capacities and through which channels. In this day and age, feedback has the potential to be a real-time, interactive, multimedia experience. Furthermore, the virtual environment inherent to elearning can ideally be set up in a fashion that simulates interaction and quality control practices that arise within the lifespan of a given professional translation project. Robinson et al. (2006) call attention to the importance of providing learners with established rubrics and modeling their utilization for optimizing self- and peer-assessment (p. 124). Technological advancement has also resulted in new avenues for assessing performance, many of which have been borrowed from the translation process research community, particularly derived from research on translation competence (Hurtado Albir, 2017). Retrospective process-oriented performance assessment can now be carried out using such technology as audio recordings of verbalized thought processes (Jääskeläinen, 2002), keystroke logging (Hansen, 2006), and screen recordings (Angelone, 2012). The latter two technologies are gaining ever-firmer footing in the self-assessment of processes thanks to their preservation of ecological validity. In gaining insightful data for purposes of retrospective reflection, the translator does not have to do anything he or she would otherwise not do while translating, such as having to articulate all thought processes or break away from the task at hand to enter content in a translation log. Eye-tracking technology holds potential in this regard as well as it continues to become less intrusive and more mobile.

Product and Process-Oriented Approaches

The range of self- and peer-assessment training approaches currently in use can be described as product-oriented or process-oriented (or a combination of both). Product-oriented approaches assess translation products as such (draft versions, final versions, revisions, etc.). This could take the form of students grading each other's translations using a specific set of assessment criteria, such as error codes and severity scales. Such tasks can be conducive to enhancing the learner's editing skills and help make rubrics more transparent (Robinson et al., 2006, p. 124). Product-oriented assessment tends to be largely quantitative

and summative in nature, with error types and frequencies, weighted point calculations (points deducted and bonus points added) to the fore. Errors and translation shifts of various kinds are often examined from the perspective of deviations from the source text as well as from comparable texts in the target language for the genre in question. As has been pointed out in the literature, product-oriented assessment tends to be dominant in translator training (Galán-Mañas & Hurtado Albir, 2015, p. 67) as well as within the language industry (Orlando, 2012, p. 205).

Process-oriented training (Gile, 1994; Lee-Jahnke, 2005; Angelone, 2014; Galán-Mañas & Hurtado Albir, 2015), though still somewhat overshadowed by product-oriented approaches, has a relatively long history, dating back to the 1980s when Gile first introduced IPDR (Integrated Problem and Decision Reporting) logs as a means by which students document problems and problem-solving strategies. Whereas product-oriented assessment focuses on errors and shifts, process-oriented assessment focuses on such phenomena as decision-making, problems (as related to, yet potentially distinct from errors), information retrieval, workflow patterns, and the manifestation of various translation competencies (Hurtado Albir, 2017). While product-oriented assessment focuses primarily on *what* has been translated, process-oriented assessment examines *how* the product came into existence in the first place. Process-oriented assessment will ideally put students in a position to “realize when they are confronted with a translation problem, verbalize it and learn how to identify it by name” and “search for an appropriate solution” (Gambier, 2012, 165).

Screen Recording as a Vehicle

While making students more cognizant of their decision-making and problem-solving tendencies for purposes of self-assessment is not a new focal point in translation (Massey & Ehrensberger-Dow, 2011a, p. 27), the potential that process-oriented self-assessment offers translator training has yet to be tapped. This holds particularly true in the domain of problem awareness training. To date, screen recording has been implemented as a diagnostic tool used by students to detect and mitigate errors in their own translations (Angelone, 2013) and for purposes of revising translations of others (Shreve, Angelone, & Lacruz, 2014). Both of these studies point towards heightened efficacy in problem recognition and subsequent error mitigation stemming from the highly visual nature of problem indicators as rendered in screen recordings. These indicators, to be discussed and contextualized in greater detail in the following section of this chapter, are likely more salient in this visual mode of self-reflection than problem indicators found in other, less visual modes of self-reflection, such as IPDR logs, think-aloud protocols, or keystroke log output. A lack of saliency when these other modes of reflection are used by students may account for observed patterns of weak problem detection and reporting (Angelone, 2010). Screen recording has also resulted in interesting problem awareness research on information retrieval tendencies (Massey and Ehrensberger-Dow, 2011b) and, more recently, human-computer interaction and cognitive ergonomics in the workplace (Ehrensberger-Dow & O’Brien, 2015). In addition to offering particularly salient rendering of problem indicators, screen recording is also advantageous in that it is free, universal in terms of operating systems, and ecologically valid in that beyond starting and stopping the recording, translators are not required to do anything they would not otherwise be doing while engaged in the task at hand (Angelone, 2015).

In extending on some of these existing approaches to fostering the translator’s problem awareness via screen recording, the approach to be outlined in this chapter will focus on how screen recordings can be utilized for retracing errors marked up in corresponding translation products. Whereas screen

recording-based training approaches posited to date have gravitated towards detection of problems for purposes of ultimately mitigating errors in revised products, the approach described here uses errors as a point of departure to be reverse engineered through screen recording diagnostics. From a pedagogical perspective, the primary objective does not lie in mitigating errors per se, but rather in being able to define and describe them once they have occurred in a more granular fashion than what is possible when using product metrics alone. Product-oriented assessment metrics commonly focus on errors at various linguistic levels, such as punctuation, grammar, lexis, syntax, style, and so on. For example, a translator who makes a lexical error might receive information regarding its linguistic level by virtue of a code such as ‘term’, ‘collocation’, ‘word choice’, or suchlike. While this information is certainly helpful, the translator is likely still left in the dark, relatively speaking, with this information alone. Is this lexical error the result of faulty internal or external support? Does the error stem from comprehension, transfer, or production problems? Did it occur during drafting or revision? These are the kinds of questions that screen recording technology, as the vehicle driving the approach to be outlined in this chapter, can help answer in providing translators with a more robust framework for classifying errors from both product and process perspectives. By virtue of engaging in such classification, in and of itself, translators and translator trainers are engaging in a form of quality assessment by discerning error patterns, but, more importantly, potential error triggers. Through cognizance of such triggers, honed through the process of classification to be described in this chapter, translators are ideally equipped with self-reflective capacities to better understand why errors have occurred, with an eye towards mitigating errors of a similar type and scope when encountered in future translation work.

This chapter will begin by contextualizing the proposed approach as optimally suited for self-regulated learning in translator training. Attention will then turn to a discussion of a range of concrete, directly observable problem indicators found in screen recordings and of how these align with various classification parameters that can be used when assessing the nature of errors found in the product. This will be followed by a description of a small-scale pilot study designed to investigate the saliency of problem indicators and inter-rater consistency when BA and MA Translation students use these parameters for classifying errors when watching the same screen recording and using the same set of marked-up errors in a translation product as a point of departure.

PROBLEM AWARENESS AT THE HEART OF SELF-REGULATED LEARNING

Educational psychologists Zimmerman and Schunk define self-regulated learning (SRL) as “the processes whereby learners personally activate and sustain cognitions, affects, and behaviors that are systematically oriented toward the attainment of personal goals” (Zimmerman & Schunk, 2011, 1). In the context of product-oriented translator training, these goals, in very broad terms, encompass mitigating errors. From a process-oriented perspective, a related goal might involve seeking an understanding of why errors occurred in the first place. For future success in the world of work, it is crucial for students to gain an awareness of how to successfully self-regulate learning of error triggers based on empirical reflection on their own performance. In fact, students need to take charge of their learning along these very lines. For all intents and purposes, it is the students themselves who are in the best position to pinpoint why a given error has potentially occurred if other assessors, such as trainers, have nothing more to go on than the translation product alone.

Interestingly, self-regulated learning has a fairly long history as a common practice in the context of interpreter training (Arumí & Esteve, 2006), where it is regarded as a form of diagnostic assessment (Choi, 2006; Postigo Pinazo, 2008). Students create audiovisual recordings of their performance and then engage in retrospective analysis in an attempt to pinpoint the cause of difficulties and, subsequently, overcome obstacles (Choi, 2006, p. 275). In line with SRL theory, students can use these recordings as process protocols or “self-oriented feedback loops through which they can monitor their effectiveness and adapt their functioning” (Zimmerman & Schunk, 2011, p. 1). In interpreter training circles, self-regulated learning has long been regarded as a bridge of sorts linking academic training with the professional world. As Lee-Jahnke (2005, p. 2) points out, “Self-assessment is not only important during the training phase of interpretation, but it is critical to professional interpreters as well. In most cases, professional interpreters are responsible for their own quality of performance and do not rely on others for assessment”. While quality assurance stages are sometimes built into the lifecycle of a translation project, with revisers and other specialists checking on translation quality, professional translators are also expected to self-assess their performance as well.

If we apply self-regulated learning approaches as used in interpreter training to the training of translators with the very same set of core objectives in mind, screen recording technology is a viable option for providing learners with access to visual evidence of real-time holistic performance. Basically, everything that happens on screen, from pauses when encountering problems, to information retrieval using online resources, to the utilization of CAT tools, to workflow patterns, is captured in an ecologically valid fashion. Students can watch translation tasks unfold in retrospect just as they occurred in real time.

Scholars in the field of educational psychology underscore the fact that self-regulated learning is particularly applicable in improving the achievement of students in a heterogeneous learning environment where great ranges in proficiency are commonplace (Zimmerman & Schunk, 2011, p. 1). This tends to be the norm in translator training, where variation at the level of variables such as translation experience, language proficiency, domain interests, and translation-relevant competencies (Hurtado Albir, 2017) often presents complex pedagogical challenges. Particularly in large enrollment translation practice courses, learners need to be in a position to take charge of their learning when it comes to problem awareness given the fact that what is problematic for some will likely not be problematic for others and vice versa. It would be virtually impossible for the trainer to address the specific learning needs of each individual student in a holistic fashion on a consistent basis throughout a given semester in such a heterogeneous environment. Marking up errors in students’ translations is a beneficial, and, more importantly, feasible starting point from the perspective of the trainer’s workload. By providing learners with concrete frameworks of what to look for in their screen recordings in conjunction with these errors, trainers can catalyze deeper learning in the domain of problem awareness through self-regulated learning. The next section of this chapter will outline what translators can be encouraged to look for in screen recordings in this capacity.

Problem Indicators as Reflected in Onscreen Activity

When watching their performance unfold in the context of a screen recording, translators can be prompted to look for a broad range of indicators that potentially signal uncertainty, which Angelone and Shreve define as “a cognitive state of indecision marked by a distinct class of behaviors” (2011, p. 108). These indicators arise when the translator faces a problem nexus. This involves the confluence of a given textual property and level with a deficiency in the translator’s declarative and/or procedural knowledge

Table 1. Problem indicator types with corresponding contextualized examples

Problem Indicator Type	Example
Extended pause (5+ seconds)	The translator is in the process of considering three possible target language equivalents for a given term and is not sure which one fits the given context, as revealed through an extended pause in on-screen activity.
External information retrieval	The translator is unsure if a given collocation is frequently used in the target language, opens a separate window/tab, and runs a search using an online field-specific corpus.
Revision	The translator is undecided about the word order in an already-generated sentence in the target text and repeatedly tweaks things until deciding on a final syntactic variant.
Mouse hovering	The translator uses the mouse to hover over a culture-specific item in the source text for which a target language equivalent does not readily come to mind.
Deviation from established routines	The translator exhibits a tendency to change the color of already-translated text to blue. A problem arises, resulting in the translator no longer highlighting newly translated content in blue.

(Angelone, 2010, p. 109). At this stage, it is important to emphasize that problems, as manifest in screen recordings, are not synonymous with errors by default. While problems do, in fact, often result in errors, these indicators of uncertainty can more broadly be conceptualized as placeholders where errors are more prone to occur and where problem-solving pathways might deviate from an optimal course of action. It is precisely these instances that translators can closely monitor when reflecting on screen recordings for the purpose of noticing, and ultimately fixing, potential errors prior to submitting a revised translation.

Perhaps the most salient problem indicator type found in screen recordings, and one that is to the fore in translation process research at large, is an extended pause in on-screen activity. While the duration of such pauses is often at the level of milliseconds for the purpose of experimental research on cognitive processes in translation, for the purpose of translator reflection in pedagogical contexts, the minimum duration is often set at a readily discernable threshold of 5+ seconds. Such pauses may or may not be followed by a second problem indicator type, namely, external information retrieval. This could, for example, take the form of using a bilingual dictionary or glossary to retrieve the target language equivalent for a given term in the source text. A third problem indicator type involves instances of revision of already-generated target text content. This revision behavior may or may not be preceded by a pause and may or may not fix an error. Several studies have pointed towards a tendency for translators to over-revise (Shih, 2006; Künzli, 2007), and, in the process, introduce errors that would otherwise not have occurred had revision not taken place.

More subtle problem indicators include phenomena such as mouse hovering over a problematic aspect of the source and/or target text and any deviation from established translation routines. Such a deviation might involve the translator turning to a translation environment resource that would not usually be used. Deviations might also take the form of changes in text mark-up patterns, as carried out, for example, to indicate problematic passages that the translator will return to at a later point in time. Table 1 provides an overview of pertinent problem indicators that are directly observable in screen recordings along with corresponding contextualized examples of each from a process standpoint.

This is by no means an exhaustive list of the problem indicators potentially found in screen recordings. Phenomena such as cursor repositioning, font changes, and misspellings, which Muñoz Martín (2009) highlights as potentially insightful indicators of attention lapses, could also be regarded as such. In this chapter, a ‘less is more’ approach is deliberately taken, centered on the indicators mentioned in Table 1 only. It is also important to point out that problem indicators may appear either in isolation or in

constellations. For example, a given problem-solving sequence could consist of the following succession of interrelated, partially overlapping problem indicators: The translator pauses when encountering a term for which he/she does not readily know the target language equivalent, opens three windows containing appropriate external resources for assistance, and hovers over various proposed solutions with the mouse before deciding on a variant to use in the target text.

The first step of the approach to reverse engineering errors described in this chapter involves having translators identify the presence and type of problem indicator(s) co-occurring on screen in conjunction with errors marked up in a corresponding translation product. The second step involves classifying the errors according to various parameters which will be described and contextualized in the next section.

EXPANDED ERROR CLASSIFICATION PARAMETERS

The past two decades of translation process research have introduced new ways of conceptualizing problems and errors in translation that complement product-oriented metrics. To date, these conceptualizations have not truly found a place in translation assessment, however. Screen recording, when used in conjunction with translation products for the purpose of more holistic assessment, may serve as an ideal bridge in this regard. One such conceptualization puts forward the notion of translation locus (Angelone, 2010). Applied to translation assessment, this involves the classification of errors according to one of the following three loci: 1) comprehension, 2) transfer, or 3) production. Comprehension errors may be based on not understanding the content of the source text, or stem from a skills deficit in the source language. Indicators of comprehension problems in screen recordings might include the translator having to look up the meaning of a given term in the source text using a monolingual tertiary resource such as a dictionary or glossary. Transfer errors are based on problematic mapping of source text content to corresponding equivalents in the target language. In such situations, the problem does not stem from gaps in knowledge of the content or source language weaknesses, but rather from either not knowing target language equivalents (declarative knowledge) or how to effectively retrieve them (procedural knowledge) (Alves, 2005). Screen recording indicators of transfer problems include the utilization of bilingual resources of various forms (dictionaries, glossaries, aligned ST-TT segments in a translation memory, etc.). Production errors are target language in orientation and stem in large part from formulation problems. This might involve a situation where the translator can readily generate three or four possible target language equivalents (hence, the error does not involve transfer), but ends up using a variant that is not adequate given the situational constraints of the translation task at hand. Indicators of production problems include monolingual information retrieval using target language resources, such as comparable texts or corpora.

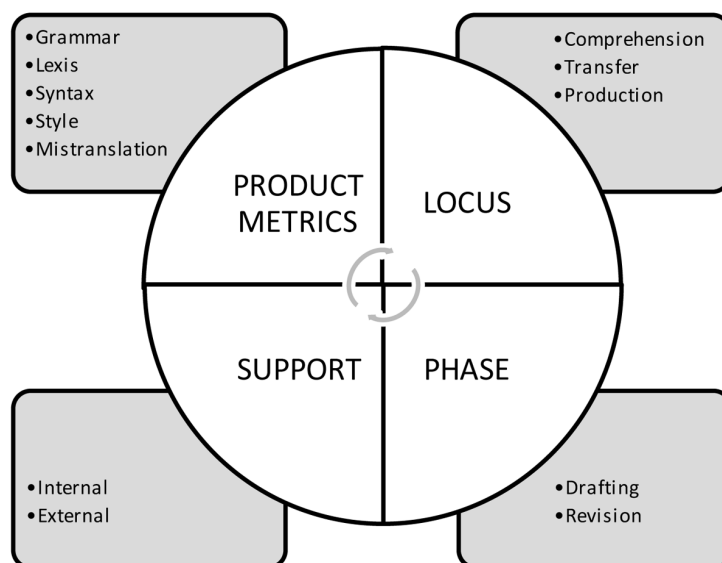
A second parameter for classifying errors that is discernable through screen recording diagnostics involves the type of support being used when the error occurred. The basic dichotomy here is internal support, where the translator taps into internalized cognitive inputs, or external support, where there is evidence of external resources being utilized for information retrieval (PACTE group, 2009). In the case of errors resulting from inefficacious utilization of external support, translators might retrieve information from unreliable resources, or perhaps the type of the resource being used does not match the linguistic nature of the problem. For example, a collocation error might result from one-word information retrieval using an online bilingual dictionary. Information retrieval involving a search for the collocation as an exact phrase multi-word unit in a parallel corpus might have prevented the error from happening.

A third relatively straightforward parameter for classifying errors involves assessors determining if they occurred during the drafting phase or the revision phase, with errors occurring during a first pass at target text generation fitting into the first category and errors occurring during revision thereof fitting into the second category. Figure 1 provides an overview of the different process-oriented error classification parameters presented in this section alongside typical product metrics.

In summary, by reverse engineering errors marked up in translation products according to the parameters of locus, support, and phase, translators can gain a more nuanced understanding of the problems they have encountered from a process-oriented perspective. This insight, enabled by screen recording technology, links the process with the product and thereby complements typically used linguistic level error classifications. The insight gained by translators and translation trainers from this triangulated error classification approach is intended to not only provide possible explanations for errors in the form of concrete triggers, but also feed into optimizing such processes as problem recognition, information retrieval patterns, and general translation workflow.

As translator training continues to move in a process-oriented direction thanks to technological advancement and concrete approaches and activities outlined in the literature, it is anticipated that there will be greater focus on implications for translation assessment. One of the ways in which screen recording-based approaches might facilitate assessment involves the potential it brings for inter-rater consistency when it comes to assigning errors to classifications thanks to the saliency of rendered problem indicators. While some componential product-oriented assessment tools have shown evidence of inter-rater consistency (Colina, 2008; Mariana, Cox, & Melby, 2015), it is still not uncommon for the same translation to be assessed by multiple assessors in an entirely different fashion, even in instances when the same grading framework is being used. This lack of inter-rater consistency is likely to become even more of an issue when severity points are used to assess translation quality, as is common in academic contexts. While grades necessitate point-based translation assessment in academic contexts to a large extent, the language industry does not face such a constraint, and, instead, takes a more dynamic approach to defining quality

Figure 1. Error classification parameters



in a much more multidimensional fashion (O'Brien, 2012). In the spirit of more holistic assessment of translation performance, a small-scale pilot study was conducted in conjunction with the approach proposed in this chapter to gauge the extent to which inter-rater consistency is obtained when BA and MA students of translation are asked to classify errors according to what they observe in a screen recording. The study design and results will be discussed in the next section.

PILOT STUDY: INTER-RATER CONSISTENCY IN ERROR CLASSIFICATION USING PROCESS DATA

As a proposed process-oriented component of an overarching holistic translation assessment paradigm, the approach outlined in this chapter was contextualized in a small-scale pilot study in an initial attempt to shed light on the following three questions:

- When watching the very same screen recording and problem indicators appearing therein, to what extent are translators consistent in their error classifications on the whole?
- Are there certain error classification parameters that yield greater or lesser consistency compared to others?
- Are there any variations in classification efficacy between the BA students and the MA students?

PARTICIPANTS

This study was conducted as an in-class activity in two German-English translation practice courses towards the end of the Spring 2017 semester at the Zurich University of Applied Sciences. The first course was a first-year BA course and 12 students took part. The second course was a first-year MA course and six students took part. To establish balance for purposes of comparison, data for only six of the twelve BA students was selected in a randomized fashion for inclusion in this study. All students in both courses were German L1 and English L2. Over the course of the semester in which the study took place, class time was dedicated to explaining problem indicators, describing error classification parameters, and whole-class practice classification activities. All students in both courses had the same level of preparation in this sense to mitigate prior exposure to the relevant concepts as a potential confounding variable. Several of the MA students graduated from the BA program, but they had not been introduced to any of these concepts as part of their undergraduate studies. In other words, the approach was equally new to both populations.

MATERIALS AND TASK

The participants were given the task of reverse engineering a marked-up translation product (see Appendix 1) containing six errors in conjunction with a corresponding screen recording by classifying each of the six errors according to 1) locus (comprehension, transfer, or production), 2) support (internal or external), 3) and phase (drafting or revision). Additionally, for each error, they were asked to identify

Table 2. Operational definitions for product metrics for purposes of error classification

Metric	Definition/Description of Errors
Grammar	Mechanical issues such as spelling, punctuation, morphology, preposition and pronoun errors
Lexis	Word choice errors, collocation errors, terminological errors that do not result in meaning being lost
Syntax	Sentence-level errors involving word order, incompleteness, and run-ons
Style	Register errors based on the defined translation brief (audience and purpose)
Mistranslation	Errors in which meaning was lost or erroneously changed as well as problematic omissions/additions

any observed problem indicators (extended pause, external look-up, revision, and/or deviation from established routines) and to also classify it according to grammar, lexis, syntax, style, or mistranslation.

The students were introduced to the information found in Table 2 in advance of the study for purposes of classifying the errors according to product metric:

The students were given a template handout for completing this classification activity (see Appendix 2). They worked on this independently and were asked not to consult with each other while doing so. Although the activity was not timed, in each course, approximately 30 minutes were spent on the classifications. All students watched the same screen recording, which was created by the instructor in an attempt to depict the types of errors from the perspective of the classification parameters in a varied fashion. A deliberate decision was made to have the translators watch the screen recording in projected format on one screen rather than on their own computers to mitigate potential technological glitches due to playback software not being installed on all student computers. That being said, ideally, in a truly pedagogical setting, this approach would lend itself well to application outside of the classroom setting, with students analyzing screen recordings (both their own and others) on their own computers in the spirit of self-regulated learning.

While the students were watching the projected screen recording, the instructor paused playback after the on-screen activity in conjunction with each error so that students would have time to complete their classifications using the provided template. The sequence of activity in conjunction with each error marked up in the product was played back twice. The screen recording the students analyzed was eight minutes in length and was an excerpt of a longer screen recording capturing the German to English translation of the mission statement for a Swiss beverage company. It was created using Screencast-o-Matic¹, a free web-based screen recorder that enables the creation of up to 15 minutes worth of video content.

Assessment of inter-rater consistency in student classifications was based on adherence to or deviation from a classification key created by the instructor. It is important to note here that this is a limited approach to inter-rater consistency, but one that is rooted in the pedagogical constellation at hand. In creating the screen recording, the instructor ensured that a relatively broad range of error classifications would align with the on-screen activity to be analyzed. Table 3 depicts how each of the six errors was classified according to the key.

Error 1 involved a lexical error in which the verb ‘gehören’, as used in the German passage ‘Rivella gehört in der Schweiz zu den stärksten Getränkeunternehmen’ was incorrectly translated as ‘Rivella belongs to the strongest beverage companies in Switzerland’. A correct verb choice would be something like ‘ranks among’. In conjunction with this error, the translator turned to a bilingual dictionary, which facilitated classification as transfer in terms of locus and external in terms of support. In addition to external information retrieval, an extended pause was also observable as a problem indicator. This error

Process-Oriented Assessment of Problems and Errors in Translation

Table 3. Error classifications key for study activity

Error	Locus	Support	Phase	Type of Problem Indicator	Product Metric
1	transfer	external	drafting	extended pause and info retrieval	lexis
2	production	internal	revision	(over)-revision	lexis
3	production	external	revision	pause, external info retrieval, deviation from established routines	mistranslation
4	comprehension	external	drafting	extended pause, external info retrieval	mistranslation
5	transfer	internal	revision	revision, deviation from established routines	mistranslation
6	production	external	revision	extended pause, external info retrieval, (over)-revision	grammar

was made while drafting rather than while revising in that corresponding target text content was not already generated at the time of occurrence.

Error 2 also involved a lexical error in which the German collocation ‘ausländische Märkte’ was translated as ‘foreign markets’ instead of ‘international markets’. Given the fact that this text is a mission statement for the beverage company, ‘international’ could be regarded as preferable in that it takes on a positive connotation and can be regarded as more global in scope than ‘foreign’. In conjunction with this error, the translator tapped into internal support, as no external resources were utilized. It can be classified as involving the locus of production in that the translator, through several iterations of revision, enters two or three potential lexical equivalents. In other words, the problem locus involved narrowing options down rather than generating them in the first place, in which case transfer would have been a more appropriate classification. Other than revision, no other problem indicators were evident on screen for this error.

Error 3 involved a mistranslation error in that meaning was altered. Here, the German passage ‘Wir machen Durstlöschen zum Erlebnis’ was mistranslated as ‘we make quenching thirst a human experience’. The notion of ‘human’ is not rendered in the source text. A better translation would have been something along the lines of ‘we make quenching thirst an unforgettable or fantastic experience’. The locus of the error is production in that the translator generated multiple options in the target text before using a monolingual corpus in an unsuccessful fashion to retrieve ‘human experience’, which would have perhaps worked well as a collocation in other contexts, but not here. The phase of the error involves revision. In terms of support, the error occurred in a situation involving external information retrieval from a monolingual (target language) corpus. Problem indicators included deviation from established routines (the translator highlighted this passage in yellow to indicate a problem, a behavior not exhibited anywhere else in the source text), an extended pause, and external information retrieval.

Error 4 was also a mistranslation error. In this case, the German passage ‘Rivellanerinnen und Rivellaner’ was mistranslated as ‘Rivellers’. This is a rather difficult construct to translate in that, given the context, it needs to be clear that these are people who work for the company, calling for the inclusion of some sort of a noun, such as ‘Rivella employees’ or ‘the Rivella team’. This important semantic dimension is lost in translation, which is why this is a mistranslation error rather than a lexical error. We can classify this error’s locus as involving comprehension in that the translator tried to retrieve definitions and descriptions for the German construct (to no avail) from bilingual dictionaries, glossaries, and by cross-checking German and English pages of the company website. The error occurred during the drafting phase and problem indicators consisted of an extended pause and external information retrieval.

Error 5 was a third mistranslation error. On this occasion, the translator translated the German passage ‘Rivallanerinnen und Rivellaner gestalten die Zukunft’ as ‘Rivellers help shape the future’. The error here lies in the inclusion of the verb ‘help’, an unnecessary and misleading word that does not appear in the source text. The correct translation would be ‘Rivella employees shape the future’. Inclusion of ‘help’ is problematic in that it undermines the capacities of these employees. In this particular case, the locus of the error is not so clear-cut and could be regarded as involving comprehension or transfer. Since the error occurred in conjunction with internal support, it is not possible to monitor the utilization of external resources as a gauge for determining if the error involves comprehension (which would involve the utilization of monolingual resources providing definitions and descriptions) or transfer (which would involve bilingual resources of various kinds). For the purposes of this study, either option (or both) was regarded as a correct classification. This error occurred during the revision phase in that the verb ‘help’ was added after the fact, i.e., after the entire sentence had been generated. Beyond revision, a second problem indicator in this instance involved deviation from established routines (in this case, the entire sentence contained more than ten misspellings that needed to be corrected, something that did not happen elsewhere in the text).

Error 6, a grammar error, occurred when the translator used an incorrect preposition in the passage, ‘shape the future over passion and determination’. The correct preposition here would be ‘with’ or ‘through’. The locus involves production in that the translator generated two options, one containing the incorrect preposition and the other containing ‘with’, but then erroneously deleted the correct alternative. The error occurred in conjunction with external support, once again involving the utilization of a monolingual corpus and comparative searches using both alternatives. In addition to revision and external information retrieval, a third problem indicator was observable in the form of an extended pause.

Students completed the exercises by watching screen recording excerpts in conjunction with errors marked up in corresponding translation products and then checking off options on a hard copy template. This template was then collected by the instructor for purposes of analysis. The students were explicitly instructed not to include their names on the sheet to preserve anonymity. They were also informed that this activity would not be formally graded and that the results would have no impact on their final course grade. They were told that the exercise would gauge their understanding of the various error classification parameters and provide them with practice in screen recording-based diagnostics. To ensure full disclosure, they were also informed that the results would be used for research purposes and that, if interested, they could have access to the results.

RESULTS AND DISCUSSION

Table 4 shows inter-rater consistency when comparing student classifications on the whole in relation to classifications found on the key. A total of 72 judgements were made, representing 12 raters classifying six errors along each parameter. No distinction is made between the classification tendencies of BA students and those of MA students at this level of analysis. Areas of correspondence and deviation are depicted according to error classification parameter type.

Given these percentages, we see that inter-rater consistency was most strongly established for the error classification parameter of phase. The students were readily able to classify errors as occurring in conjunction with drafting or revision. The relatively straightforward dividing line is whether or not target text content had already been generated. This distinction proved to be particularly salient, as might be

Process-Oriented Assessment of Problems and Errors in Translation

Table 4. Inter-rater consistency between student classifications and key classifications

Error Classification Parameter	% Correspondence	% Disagreement
locus	52/72 = 72%	20/72 = 28%
support	47/72 = 65%	25/72 = 35%
phase	64/72 = 89%	8/72 = 11%
product metric	59/72 = 82%	13/72 = 18%

expected from a purely visual standpoint. Interestingly, the obtained data also point toward efficacy in classifying errors according to product metric. This finding suggests that process diagnostics might serve as a vehicle for enhancing inter-rater consistency when product metrics are utilized to assess translation performance. To better gauge this potential benefit, it would be important to run an empirical study in which translation products are assessed by multiple graders in two conditions, the first involving the utilization of screen recording when marking up errors and the second without. This would seem to be a logical extension to the exploratory study presented here and further establish a potential process-product link in the domain of translation assessment.

A 72% correspondence rate was found when students classified errors according to locus. A closer examination of classifications in conjunction with each locus reveals that consistency was relatively high when transfer was involved and considerably lower when production was involved. This suggests that indicators of transfer problems (utilizing a wide array of bilingual or multilingual resources) might be more salient than indicators of production problems (utilizing an array of resources in the target language for formulation purposes). This may also point to some gray area in attempts to establish concrete boundaries between transfer and production as translation loci. Perhaps students were inclined to think that resources like monolingual corpora and target language parallel texts were being used to facilitate transfer instead of production. Without a parallel stream of data, such as articulations in a think aloud protocol or information presented in the context of a retrospective interview, there is a certain degree of guessing inherent to the classifications. The translation process research community is aware of this and, in response, has called for methods involving triangulation since the turn of the millennium (Alves, 2003; Saldhana & O'Brien, 2013, p. 23). While triangulation, i.e. having overlapping, interrelated data points, is undeniably warranted in situations of experimental research, it may be less feasible in pedagogical contexts, where time is of the essence and where students, as self-regulated learners, likely do not have the mindset or objectives of an experimental researcher. A seeming loss of rigor in design and implementation when using technology and models borrowed from TPR in academic contexts may thus be a necessary trade-off.

Interestingly, the weakest level of consistency was found for the error classification parameter of support. Apparently, the distinction between using and not using external resources for purposes of problem-solving was not as clear for the students as one might have anticipated given the basic nature of this dichotomy. Perhaps the type of support being used in conjunction with an error is not as salient an indicator as the phase at which the error occurs. Another plausible explanation is that the rater's cognitive load allocated for purposes of error parameter classification has a limit and that the parameter of support is either consciously or subconsciously deprioritized in relation to the other parameters. This tendency to prioritize and deprioritize error classification parameters could be ascertained through a retrospective interview or survey, but is beyond the scope of this chapter.

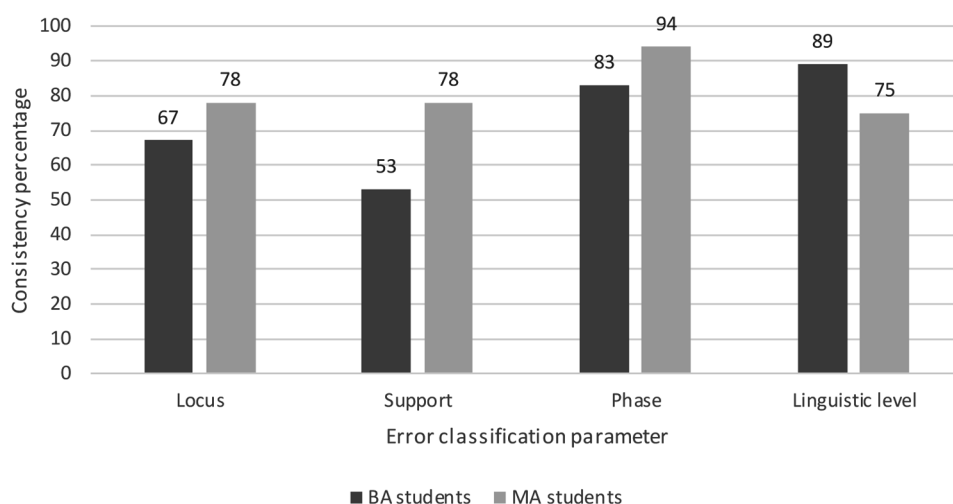
Figure 2 breaks down these results to depict inter-rater consistency at the level of the two student populations to document any potential variation depending on level of studies.

Figure 2 shows that, on average, the MA students tended to show more correspondence with the key classifications than the BA students, perhaps in large part by dint of their having had more translation practice courses. However, it is important to note that the concepts of problem indicator, locus, support, and phase, as fundamental dimensions of translation process research, were equally new to both groups of students. This suggests that the MA students likely have a more solid awareness of translation as a process to the extent that the new metalanguage being attached to it by virtue of these classification parameters is more readily grasped and used.

The largest discrepancy in consistency between the two student populations can be seen in the error classification parameter of support. Just over half of the BA students were consistent with the key in their classification of errors as occurring in conjunction with either internal vs. external support. From a purely observational standpoint, either using or not using external resources would seem to be clear-cut. However, from a conceptual standpoint, the BA students might not see a clear distinction given the fact that student translators often resort to external support by default when encountering problems. The idea of turning to cognitive inputs, such as world knowledge and context (Lederer, 2007), as internal support is likely still a more latent concept for the undergraduates. Several of the MA students had taken coursework in sight translation during the later stages of their BA degree. In these courses, the skill of deverbization was to the fore, and it is likely that they become more cognizant of the very notion and benefits of internal support in this course. At the time of their participation in this study, the BA students had yet to take these courses.

Perhaps the most interesting finding at this level of comparative analysis is the fact that the BA students showed greater consistency than the MA students when classifying errors according to product metrics. This was the only classification parameter where this was the case. The result is surprising in that the MA students had all taken several translation practice courses in which they received feedback containing these product metrics, whereas the BA population consisted of students taking their first translation practice course and they had been introduced to these product metrics as errors over the course of only

Figure 2. Inter-rater consistency according to student population in relation to a key



one semester. These results suggest a potential benefit in introducing this approach to process-oriented training and self-assessment early on in the curriculum, before these linguistic levels are solidified and seen through a product-oriented lens alone, as potentially was the case for the MA students in this study.

CONCLUSION AND FUTURE DIRECTIONS

The process-oriented approach to translation assessment outlined in the chapter aims at fostering a deeper causal understanding of errors that transcends the metrics typically used in product-oriented assessment, which is still dominant in translator training. This understanding comes from the translator being able to reverse engineer errors marked up in translation products and, ideally, classify each according to a given locus, type of support used, and phase, in addition to linguistic level. In linking the product with the process in this regard, such triangulated assessment yields an understanding of both *what* errors occurred in translation as well as *why*. The user-friendly nature of screen recording and the saliency it offers in the recognition of problem indicators facilitates problem awareness in an ideal fashion for self-regulated learning and self-assessment both in the classroom and beyond.

The small-scale pilot study discussed in this chapter suggests relative efficacy in terms of inter-rater consistency when BA and MA students use this approach for purposes of classifying errors according to the various parameters, further attesting to the saliency of problem indicators in screen recording. The BA students, in particular, exhibited an enhanced capacity for classifying and assessing errors according to product metrics by virtue of watching processes unfold on screen. It is important to emphasize that the obtained results are preliminary. At least to some extent, the documented variation between the two student populations might have resulted from the relatively small group sizes. Nevertheless, these findings would seem to warrant more expansive empirical investigation involving larger participant groups, additional language pairs, and assessment for statistical significance to see if inter-rater consistency patterns hold true in more robust contexts. The proposed process-oriented approach is envisaged as one that could be a fundamental component of a holistic approach to translation assessment at large, complementing extant product-oriented models. It is hoped that the inter-rater consistency found when students classified errors using screen recordings in the context of this study will motivate future empirical research on the applicability of such an approach in other aspects of translation assessment. For example, it might be interesting to see if a given group of assessors reaches a higher level of inter-rater consistency when assessing translations using screen recordings than when doing so without.

REFERENCES

- Alves, F. (Ed.). (2003). *Triangulating Translation: Perspectives in Process Oriented Research*. Amsterdam: John Benjamins. doi:10.1075/btl.45
- Alves, F. (2005). Bridging the Gap Between Declarative and Procedural Knowledge in the Training of Translators: Meta-Reflection Under Scrutiny. *Meta*, 50(4). Retrieved from <http://www.erudit.org/revue/meta/2005/v50/n4/019861ar.pdf>
- Angelone, E. (2010). Uncertainty, Uncertainty Management, and Metacognitive Problem Solving in the Translation Task. In G. Shreve & E. Angelone (Eds.), *Translation and Cognition* (pp. 17–40). Amsterdam: John Benjamins. doi:10.1075/ata.xv.03ang

- Angelone, E. (2012). The Place of Screen Recording in Process-Oriented Translator Training. *RITT*, 14, 41–55.
- Angelone, E. (2013). The impact of process protocol self-analysis on errors in the translation product. *Translation and Interpreting Studies*, 8(2), 253–271. doi:10.1075/tis.8.2.07ang
- Angelone, E. (2014). A Corpus-Based Comparison of Self- Reflection Modalities in Process-Oriented Translator Training. In Y. Cui & W. Zhao (Eds.), *Teaching Language Translation and Interpretation: Methods, Theories, and Trends* (pp. 346–361). Hershey, PA: IGI Global.
- Angelone, E. (2015). Optimizing Process-oriented Translator Training Using Freeware. In P. Sandrini & M. García González (Eds.), *Translation and Openness* (pp. 133–145). Innsbruck: University of Innsbruck Press. doi:10.15203/2936-88-2-7
- Angelone, E., & Shreve, G. (2011). Uncertainty Management, Metacognitive Bundling in Problem-Solving, and Translation Quality. In S. O'Brien (Ed.), *Cognitive Explorations of Translation* (pp. 108–130). London: Continuum.
- Arumí, M., & Esteve, O. (2006). Using instruments aimed at self-regulation in the consecutive interpreting classroom. *Electronic Journal of Foreign Language Teaching*, 3(2), 158–189.
- Beeby, A., Fernández, M., Fox, O., Albir, A., Kozlova, I., Kuznik, A., ... Wimmer, S. (2009). Results of the validation of the PACTE translation competence model: Acceptability and decision making. *Across Languages and Cultures*, 10(2), 207–230. doi:10.1556/Acr.10.2009.2.3
- Choi, J. Y. (2006). Metacognitive evaluation method in consecutive interpretation for novice learners. *Meta*, 51(2), 273–283. doi:10.7202/013256ar
- Colina, S. (2008). Translation Quality Evaluation: Empirical Evidence for a Functionalist Approach. *The Translator*, 14(1), 97–134. doi:10.1080/13556509.2008.10799251
- Ehrensberger-Dow, M., & O'Brien, S. (2015). Ergonomics of the translation workplace: Potential for cognitive friction. *Translation Spaces*, 4(1), 98–118. doi:10.1075/ts.4.1.05ehr
- Ericsson, K. A. (1996). The Acquisition of Expert Performance: An Introduction to Some of the Issues. In K. A. Ericsson (Ed.), *The Acquisition of Expert Performance in the Arts and Sciences, Sports and Games* (pp. 1–50). Mahwah, NJ: Lawrence Erlbaum Associates.
- Galán-Mañas, A., & Hurtado Albir, A. (2015). Competence assessment procedures in translator training. *The Interpreter and Translator Trainer*, 9(1), 63–82. doi:10.1080/1750399X.2015.1010358
- Gambier, Y. (2012). Teaching translation / Training translators. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of Translation Studies* (Vol. 3, pp. 163–171). Amsterdam: John Benjamins. doi:10.1075/hts.3.tea1
- Gile, D. (1994). The Process-Oriented Approach in Translation Training. In C. Dollerup & A. Lindegaard (Eds.), *Teaching Translation and Interpreting 2: Insights, Aims, Visions* (pp. 107–112). Amsterdam: John Benjamins. doi:10.1075/btl.5.17gil
- Hansen, G. (2006). Retrospection Methods in Translator Training and Translation Research. *Journal of Specialized Translation*, 5, 2–40.

Process-Oriented Assessment of Problems and Errors in Translation

Hurtado Albir, A. (Ed.). (2017). *Researching Translation Competence by PACTE Group*. Amsterdam: John Benjamins. doi:10.1075/btl.127

Hurtado Albir, A., & Taylor, P. (2015). The Acquisition of Translation Competence. Competences, Tasks and Assessment in Translator Training. *Meta*, 60(2), 256–280. doi:10.7202/1032857ar

ISO. (2015). *Translation Services – Requirements for translation services*. ISO 17100:2015. Geneva: ISO. Retrieved from <https://www.iso.org/standard/59149.html>

Jääskeläinen, R. (2002). Think-aloud protocol studies. *Target*, 14(1), 107–136. doi:10.1075/target.14.1.05jaa

Kiraly, D. (2003). From Teacher-Centred to Learning-Centred Classrooms in Translator Education: Control, Chaos or Collaboration? In A. Pym, C. Fallada, J. R. Biau, & J. Orenstein (Eds.), *Innovation and E-Learning in Translator Training: Reports on Online Symposia* (pp. 27–31). Tarragona: Universitat Rovira I Virgili.

Künzli, A. (2007). The ethical dimension of translation revision. An empirical study. *The Journal of Specialised Translation*, 8, 42–56.

Lederer, M. (2007). Can Theory Help Interpreter and Translator Trainers and Trainees? *The Interpreter and Translator Trainer*, 1(1), 15–36. doi:10.1080/1750399X.2007.10798748

Lee-Jahnke, H. (2005). New Cognitive approaches in Process-Oriented Translation Training. *Meta*, 50(2), 359–378. doi:10.7202/010942ar

Mariana, V., Cox, T., & Melby, A. (2015). The Multidimensional Quality Metrics (MQM) Framework: A new framework for translation quality assessment. *The Journal of Specialised Translation*, 23, 137–161.

Massey, G., & Ehrensberger-Dow, M. (2011a). Commenting on translation: Implications for translator training. *Journal of Specialised Translation*, 16, 26–41.

Massey, G., & Ehrensberger-Dow, M. (2011b). Investigating information literacy: A growing priority in translation studies. *Across Languages and Cultures*, 12(2), 193–211. doi:10.1556/Acr.12.2011.2.4

Muñoz Martín, R. (2009). Typos & Co. In S. Göpferich, A. Jakobsen, & I. Mees (Eds.), *Behind the Mind: Methods, Models and Results in Translation Process Research* (pp. 167–191). Copenhagen: Samfundslitteratur.

O'Brien, S. (2012). Towards a Dynamic Quality Evaluation Model for Translation. *The Journal of Specialised Translation*, 17, 55–77.

Orlando, M. (2012). Training Professional Translators in Australia: Process-oriented and Product-oriented Evaluation Approaches. In S. Hubscher-Davidson & M. Borodo (Eds.), *Global Trends in Translator and Interpreter Training* (pp. 197–216). London: Continuum.

Postigo Pinazo, E. (2008). Self-Assessment in Teaching Interpreting. *TTR*, 21(1), 173–209. doi:10.7202/029690ar

Robinson, B., López Rodríguez, C., & Tercedor Sánchez, M. (2006). Self-assessment in translator training. *Perspectives. Studies in Translatology*, 14(2), 115–138. doi:10.1080/09076760608669025

Saldhana, G., & O'Brien, S. (Eds.). (2013). *Research Methodologies in Translation Studies*. London: Routledge.

Shih, C. (2006). Revision from translators' point of view. An interview study. *Target*, 18(2), 295–312. doi:10.1075/target.18.2.05shi

Shreve, G., Angelone, E., & Lacruz, I. (2014). Efficacy of Screen Recording in the Other-Revision of Translations: Episodic Memory and Event Models. *MONTI*, 1, 225–245. doi:10.6035/MonTI.2014.ne1.7

Zimmerman, B., & Schunk, D. (2011). Self-Regulated Learning and Performance: An Introduction and an Overview. In B. Zimmerman & D. Schunk (Eds.), *Handbook of Self-Regulation of Learning and Performance* (pp. 1–12). New York: Taylor & Francis.

ENDNOTE

- ¹ <https://screencast-o-matic.com/> For an overview of additional free and/or open source screen recording applications, see Angelone 2015.

APPENDIX 1

Source Text* and Translation With Marked-Up Errors

Unsere Werte

Rivella gehört in der Schweiz zu den am stärksten, erfolgreich wachsenden Getränkeunternehmen und etabliert sich in ausgewählten ausländischen Märkten.

Was den Konsumenten bewegt, inspiriert uns: Wir machen Durstlöschen zum Erlebnis.

Rivellanerinnen und Rivellaner gestalten die Zukunft mit Leidenschaft und Siegeswillen.

* This excerpt was part of the Rivella mission statement that appeared on the company's website in October 2016. This mission statement has since been updated:

<http://www.rivella-group.com/de/unternehmen/werte>

Our Values

Rivella belongs to **[[ERROR 1]]** the strongest, successful growing beverage companies in Switzerland and is establishing itself in select foreign **[[ERROR 2]]** markets.

We are inspired by the things that move our customers. We make thirst-quenching a human experience **[[ERROR 3]]**.

Rivellers **[[ERROR 4]]** help **[[ERROR 5]]** shape the future over **[[ERROR 6]]** passion and determination.

APPENDIX 2

Template for Analysis Activity

Please classify the six errors according to the following parameters based on what you observe in the screen recording.

Error 1

Which type of support was being used when the error occurred:

___ Internal support ___ External support

What is the locus of this error:

___ Comprehension ___ Transfer ___ Production

Process-Oriented Assessment of Problems and Errors in Translation

Which problem indicators were observable (check all that apply):

None Extended pause of 5+ seconds External information retrieval
 Revision Deviation from established routines

How would you classify this error?

Grammar Lexis Syntax Style Mistranslation

[Template repeated for errors 2 to 6.]

Chapter 9

Exploring Process– Oriented Translation Competence Assessment: Rationale, Necessity, and Potential From a Problem–Solving Perspective

Si Cheng

The University of Manchester, UK

ABSTRACT

In translation education and training, competence assessment plays an essential role in fostering students' competence development. This chapter promotes process-oriented assessment from a translation problem-solving perspective, with supporting evidence from a longitudinal study. Although process-oriented competence assessment is increasingly becoming more important in translation education and training, the relation between the development of translation competence and the translation process remains underexplored. This chapter provides a translation problem-solving perspective to understand the rationale for and the necessity of process-oriented competence assessment and suggests practical and feasible process-oriented assessment tools in the translation classroom.

INTRODUCTION

Pedagogical translation assessment mainly consists of the evaluation of the learners' translation competence (Martínez Melis & Hurtado Albir, 2001). In traditional assessment approaches, the translation product has been regarded as the main object of assessment. Relevant discussions have mostly focused on developing rating scales that are based on translation errors and translation problems (e.g. House, 1997; Kim, 2009; Kussmaul, 1995). The development of translation competence is then inferred from the assessment of the product, i.e. the translation. However, as a result of a recent paradigm shift from teacher-centred transmissionist to learner-centred constructivist approaches (Kim, 2013), a number of translation scholars have begun to advocate for process-oriented approaches (e.g., Angelone, 2016; Gile,

DOI: 10.4018/978-1-5225-5225-3.ch009

2009; Kiraly, 2000; Massey, 2005; Orlando, 2012; Zhong, 2005). Consequently, assessing the learners' development of translation competence has also become increasingly important.

Process-oriented pedagogical approaches suggest the use of some tools to gather more detailed information about translation processes. For example, Fox (2000) uses translation diaries in a process-oriented competency¹-based translation course. The author encourages students to keep a record of the translation process, including how they solve translation problems and the rationale behind their translation choices. Similarly, Gile (2004) proposes the adoption of integrated problem and decision reporting as a pedagogical tool. To this end, students are required to report on the problems they encounter during the translation process, the steps they take to solve them and the rationale for their final decisions. In Shei's (2005) translation classroom, the students are required to analyse the text and the context, explain the problem-solving procedures and record other task-related thoughts in their translation commentaries. Orlando (2012) also encourages the use of a translator's diary as a pedagogical tool for instructors to have a better understanding of the translation processes followed by students and to discover adequate teaching and remedial strategies. However, the relevance of these pedagogical tools to the development of translation competence is still underexplored, with these discussions mostly focusing on the novelty or effectiveness of such pedagogical tools. In other words, further research is necessary on how to assess the development of translation competence as reflected by the translation process.

Therefore, this chapter aims to introduce a problem-solving perspective to conceptualise the development of translation competence as part of the translation process, which provides a rationale for process-oriented approaches to the assessment of translation competence. The proposed conceptualisation is also supported by empirical evidence from a longitudinal study with a small group of translation students. Findings from this empirical study highlight the necessity of process-oriented assessment and suggest a more effective use of relevant pedagogical tools in translation education and training.

THE CONCEPTUALISATION OF TRANSLATION COMPETENCE

As the development of translation competence is the main object of study in pedagogical assessment, the conceptualisation of translation competence and its development is essential since it lays the foundation for understanding and implementing effective forms of assessment in translator education (Huertas Barros and Vine, 2018a, 2018b). This section provides a brief literature review on the conceptualisation of translation competence, with an emphasis on the translation process. It should be noted that there is a lack of terminology consistency when referring to this notion, with some authors using the term *translator competence* to entail a wider range of skills related to the communities of professional translators (see discussions in Biel, 2011; Kiraly, 2000; Pym, 2011). However, the term *translation competence* will be used throughout this chapter because the emphasis here is not on specific competence components.

The Prevalence of Multicomponential Models

For the past few decades, the conceptualisation of translation competence has been characterised by the prevalence of multicomponential models and an increase in the number and scope of competence components. In its first conceptualisations (e.g. Wilss, 1976), translation competence is seen as a supercompetence that requires interlingual competence as its prerequisite. However, more recent conceptualisations conceive it as a multicomponential concept that goes far beyond language competence.

Exploring Process-Oriented Translation Competence Assessment

From this perspective, translation competence is believed to be ‘a hierarchical configuration of clearly distinguishable component competences’ (Neubert, 1994, p. 419). The identification of these competence components, as Schäffner and Adab (2000) highlight, is considered essential for translation competence research. Various attempts have been made over the years to define the different components of translation competence, with the emergence of new components highlighted by different models, e.g. communicative competence (Bell, 1991), disposition (Campbell, 1991), research competence (Nord, 1992), subject competence (Neubert, 1994), monitoring competence (Campbell, 1998), instrumental competence (PACTE, 2003), interpersonal competence (Kelly, 2007), translation routine activation competence (Göpferich, 2009), translation service provision competence (EMT, 2009), and ethics competence (NAATI, 2015).

As Pym (2003) points out, a major purpose for the expansion of multicomponential models is ‘to bring new skills and proficiencies into the field of translator training’ (p. 481). To this end, translation trainers and scholars have proposed various multicomponential models to enhance curriculum design. Thus, the various components that characterise these models provide a framework for setting learning objectives, designing class activities, or assessing learning outcomes. For example, Fox (2000) specifies five abilities as the learning goals for her Catalan-English translation course, which reflect the following five translation competencies: communicative competence, sociocultural competence, language and cultural awareness, learning how-to-learn and problem-solving goals. Kelly (2007) proposes one of the most influential translation competence models for curriculum design, which includes the following seven areas of competence: communicative and textual, cultural and intercultural, subject area, professional and instrumental, attitudinal or psycho-physiological, interpersonal or social, and organisational or strategic competence.

Some translation researchers have also proposed multicomponential models to serve as a conceptual basis for empirical studies on the development of translation competence. The PACTE group, for example, builds its empirical project upon their proposed translation competence model (PACTE, 2003) which comprises five subcomponents: bilingual, extralinguistic, knowledge about translation, instrumental, and strategic subcompetence. In addition, translation competence also activates a series of psycho-physiological mechanisms. Consequently, its acquisition is a process in which the subcompetences forming the overall translation competence are ‘developed and restructured’ (PACTE, 2017, p. 304). Göpferich (2009) proposes a similar translation competence model for the TransComp project, a longitudinal research project on the development of translation competence. Göpferich’s model differentiates between six subcompetences: Communicative competence in at least two languages, domain competences, tools and research competence, translation routine activation competence, psychomotor competence, and strategic competence. Drawing from dynamic systems theory, Göpferich (2013) further argues that the subcompetences interact over time during the process of competence development, and that different subcompetences may not develop at the same pace or in a linear manner.

Furthermore, multicomponential models for qualification and accreditation purposes have been put forward by certain relevant institutions or committees. The European Master’s in Translation (EMT) expert group (2009) has developed a competence framework as a reference for member institutions of the EMT network, considering a variety of research projects, professional literature and professional experience. The EMT model includes six areas of competence: translation service provision, language, intercultural, information mining, technological, and thematic competence. Under each area, relevant knowledge and skills are also specified. The National Accreditation Authority for Translators and Interpreters (NAATI) has also made an attempt to map individual skills and knowledge areas from the

Australian Government's Public Sector Training Package and match them against existing competence frameworks. As a result, the NAATI has identified eight common aspects of translation competence: Language, intercultural, research, technological, thematic, transfer, service provision and ethical competency (NAATI, 2015). The NAATI's proposal provides a detailed list of knowledge and skills under each competence area as well as relevant attributes.

The Interrelation Between Competence Components

The prevalence of these multicomponential models of translation competence is considered beneficial by some scholars. Indeed, according to Kelly (2007), the compartmentalisation of translation competence into different areas has the advantage of dividing the overall competence into manageable and easily assessable units. This conceptualisation of translation competence provides a solid basis for a flexible design of the translation curriculum, and facilitates research on the development of each component including research on how to assess each individual component. However, these multicomponential models also suffer from an intrinsic problem: the interrelation between the different components in the translation process and in the development of translation competence is unclear. As Neubert (1994) pointed out more than two decades ago, a crucial question for translation research is how these component competences 'interrelate efficiently, effectively, and adequately' to form the overall translation competence (p. 412). Kelly (2007) also agrees that some students may find it difficult to establish and understand the relations between the different components promoted by modularised approaches to translation competence.

In fact, in existing multicomponential models, some attention has already been paid to the interrelation and interaction of the subcompetencies. For example, in the PACTE model, the strategic subcompetence is described as responsible for 'activating the different subcompetences' in the translation process (PACTE, 2005, p. 610). In Göpferich's (2009) model, the strategic competence is regarded as a metacognitive competence that 'sets priorities and defines hierarchies between the individual subcompetences' (p. 23). The NAATI (2015) also stresses that the competences do not exist in isolation and that, when completing a translation task, 'a translator will integrate many or all of these areas of competency in order to accurately translate a source text into a target text' (p. 8). Similar discussions can also be found in Prieto Ramos (2011) and Way (2014). However, these debates rarely go beyond a mere description to explain how strategic competence activates, sets priorities, and defines hierarchies between the various subcompetences and how a translator integrates the different competence components during the translation process. In other words, there is a need for further research on the interrelation and interaction between the different components that comprise translator competence. Consequently, a further unexplored area is how the interrelation between these components changes in the development of translation competence.

Process-Oriented Approaches to Translation Competence

Some translation scholars highlight the limitations of multicomponential models, and have adopted a different approach to conceptualising translation competence with more consideration of the translation process. For instance, considering that translation is a process of generation and selection between alternative texts, Pym (1992, p. 281) proposes a minimalist definition of translation competence that involves two skills: the ability to generate a target text series of more than one viable term for a source text, and the ability to select only one target text from this series quickly and with justified confidence. According to Pym, multicomponential models often operate as 'a political defence of a certain model

of translator training', and the scope of the competence components is being so overtly expanded that almost 'any number of neighbouring disciplines can be drawn on, any number of things can be included under the label of translation competence' (Pym, 2003, p. 487). While acknowledging that the translator needs to have a fair amount of knowledge, skills, and abilities in linguistic, commercial and other areas, Pym (2003) believes that 'the specifically translational part of their practice is strictly neither linguistic nor solely commercial', and translation competence is predominantly a process of 'generation and selection' (p. 487) between alternative texts.

Kiraly (2013) is also critical of multicomponential competence models, but unlike Pym's simplified minimalist view, Kiraly's model gives prominence to the complexity involved in translation processes, translation competence and translation competence acquisition. According to Kiraly (2013), a striking feature of the multicomponential competence models is that 'none of them suggests or reveals anything at all about the learning process' (p. 201). As a matter of fact, findings from expertise studies (e.g. Dreyfus & Dreyfus, 1986) reveal that competence tends to be relatively conscious and rule-based at lower levels of proficiency, and more intuitive and less rule-based at higher levels of proficiency. Therefore, Kiraly proposes adding a third dimension to the two-dimensional multicomponential models to depict the networks between the subcompetences. In other words, the interplay between the subcompetences is more conscious and rule-based in novice translators but will become more intuitive and less rule-based in professional translators. The development of translation competence is therefore reflected in the changing pattern of the interplay between the subcompetences.

Pym's (2003) minimalist definition of translation competence may seem to drift away from the multicomponential models, however Kelly (2007) highlights that Pym's conceptualisation implicitly requires the translator to possess the components that comprise multicomponential models in order to be viable. In other words, in Pym's minimalist definition of translation competence, the multicomponential models still exist, but in a different form – as an implicit prerequisite. In the view of Kiraly and Piotrowska (2014), however, competence is an emergent phenomenon – it is a 'holistic and autopoietic (dynamic, unpredictable, self-generating, and self-maintaining)' process rather than a discernable product (p. 3). It is due to existing institutional structures and widespread reductionist pedagogical beliefs that the multicomponential dimension of translation competence is preserved in the proposed model. In the process-oriented approaches to translation competence of both Pym and Kiraly, the interrelation and interaction between the subcompetences remain under-discussed: Pym has simply disregarded the multicomponential models; while Kiraly has only tentatively retained the multicomponential models and has thus given little detailed discussion on the interplay between the components.

A PROBLEM-SOLVING PERSPECTIVE ON TRANSLATION COMPETENCE

In order to address the lack of research on the interplay between the multicomponents of translator competence, this chapter proposes a problem-solving perspective to shed some light on the interrelation between the competence components and the development of translation competence during the translation process.

In translation studies, the translation process has long been associated with problem solving. Early in 1967, Levý (2000) emphasised that 'translating is a decision process' (p. 148), while Gaddis Rose (1979) argued that translation is problem solving, and '[t]he overriding problem is to get a text from one

language to another' (p. 20). In line with this approach, Tirkkonen-Condit (2000) proposes that translation is 'a goal-oriented action and can therefore be described as a problem-solving' (p. 125). This notion is also supported by Sirén and Hakkarainen (2002), who believe that translation should be conceived of as a problem-solving activity, and that from this perspective, the problem should be defined as the entire translation task, instead of single textual elements or contextual issues. Darwish (2008) places translation within an overall framework of communication and also regards translation as 'a problem-solving process that involves a wide spectrum of decisions producing a translation-mediated solution to a communication problem' (pp. 17-18). Other scholars also see problem-solving as part of the translation process. For example, Krings (1986) believes that translation strategies, the 'potentially conscious plans for solving a translation problem,' (...) 'emerge as soon as the translation cannot be carried out automatically' (p. 268). Angelone (2010) sees translation as 'a (more or less) linear progression of problem solving task sequences interspersed with unproblematic translation sequences where comprehension, transfer, and production occur unimpeded by any difficulties' (p. 18).

Problem solving is also referred to in the existing literature focusing on translation competence. For example, Wilss (1996) believes that the 'accountability of translator performance depends, first of all, on the problem-solving capabilities of the translator' (p. 47), while Shreve (2002) points out that, from a problem-oriented view of translation, the development of translation expertise could be seen as 'an increased capacity to recognise and represent the problems of translation and an increased ability to effectively resolve those problems' (p. 161). Meanwhile, Pym's (2003) minimalist definition of translation competence also builds upon the notion that translation is 'a process of generation and selection, a problem-solving process that often occurs with apparent automatism' (p. 489). Similarly, Hansen (2010) argues that translation competence means that 'problems and errors during the reception and formulation phase are recognized by the translator and that he/she has problem solving and revision strategies available to solve the problems and improve the text' (p. 191). Way (2014) also contends that the 'framework of decision making, which brings into play different subcompetences to solve different problems' is central to the translation process (p. 137). In the NAATI's (2015) list of knowledge, skills, and attributes (KSAs), a problem-solving attribute is described as being 'able to identify and find effective solutions for issues in order to achieve a goal' (p. 14). The PACTE group also believes that 'translation problems can be located in the various phases of the translation process' and solving translation problems 'involves different cognitive operations within the translation process, and requires constant decision-making on the part of the translator' (Hurtado Albir, 2017, p. 10). Problem solving is also related to competence components in many other conceptualisations of translation competence (EMT, 2009; Göpferich, 2009; Kelly, 2007; Kiraly, 1995; Prieto Ramos, 2014).

While problem solving is widely believed to be closely linked with the translation process, there are divergences over the extent to which translation can be regarded as a problem-solving process. Some scholars view translation as entirely a problem-solving process (Gaddis Rose, 1979; Levý, 2000; Pym, 2003; Shreve, 2002; Sirén & Hakkarainen, 2002; Tirkkonen-Condit, 2000); others believe that only when automatised solutions are not found and nonautomatised processes are carried out will the translation process become a problem-solving one (Angelone, 2010; Bell, 1998; Hurtado Albir, 2017; Kiraly, 1995; Krings, 1986). This chapter considers the translation process as entirely a problem-solving process for the following reasons: first, the act of translating is a problem, from the problem-solving perspective, and second, those seemingly automatised translation processes are in fact also accelerated problem-solving processes. In problem-solving studies, a problem is defined as a gap between the current state and the desired state. Hayes (1989) points out that 'whenever there is a gap between where you are now

and where you want to be, and you don't know how to find a way to cross that gap, you have a problem' (p. xii). Similarly, Robertson (2001) claims that 'when a goal is blocked, you have a problem' (p. 227). In a translation context, the current state is the source text and the desired state is the target text. Since translation is a highly contextualised activity, and the desired state, i.e. the target text, is subject to the specific translation situation, how to get from the source text to the target text is not immediately known to the translator. The translation task itself is, therefore, a problem and the process of carrying out the translation is a problem-solving process. This macro problem-solving process consists of numerous micro problem-solving processes of specific translation problems. Some translation problems may take more time and effort to solve, while other problems may be solved fairly easily so that the problem-solving processes seem to be automatised. When part of a translation process is described as automatised, it usually means that spontaneous and immediate solutions for translation problems are generated (Hurtado Albir, 2017), but these solutions are not necessarily free from conscious (and therefore unautomatised) evaluating or decision making. Apparent automaticity should therefore be regarded as a partial acceleration of the problem-solving process.

Since the translation process is perceived as a problem-solving process, what is demonstrated from this process is therefore the translation problem-solving ability. More specifically, it refers to a procedural ability that enables a translator to identify and solve translation problems during the translation process. It should be noted that in this conceptualisation, translation problems are not limited to linguistic, textual or cultural issues, but also include extra textual, technological, interpersonal, and other problems that may arise in the translation process. Identifying and solving these problems requires a combination of knowledge, skills, and attributes, which are *orchestrated* by the translation problem-solving ability during the translation process. In other words, in order to be able to translate, a translator not only needs to possess prerequisite knowledge, skills, and attributes, but also the ability to organise, integrate, and activate them while carrying out a translation task. Consequently, translation competence can be defined as a demonstrated *ability to translate* that results from orchestrating a combination of knowledge, skills, and attributes in carrying out a given translation task under given conditions (the *ability to orchestrate* in this definition is represented by the translation problem-solving ability). Therefore, the development of translation competence involves two interacting processes: the development of translation-relevant knowledge, skills, and attributes, and of translation problem-solving ability.

Given that translation problem-solving ability is conceived as a procedural ability, it is necessary to understand the different steps involved in this procedure. Pretz, Naples, and Sternberg (2003, pp.3-4) conceive problem-solving as a process consisting of the following seven steps: 1. recognising or identifying the problem; 2. defining and representing the problem; 3. developing a solution strategy; 4. organising knowledge about the problem; 5. allocating mental and physical resources for solving the problem; 6. monitoring progress towards the goal; and 7. evaluating the solution for accuracy (Other problem-solving models can also be found in e.g. Hayes 1989; Newell and Simon, 1972). Drawing from problem-solving conceptualisations, several attempts have also been made in translation studies to describe the problem-solving process. For example, Wilss (1996, p. 188) proposes a six-stage model for decision making in translation: 1. identifying the problem; 2. clarifying or describing the problem; 3. researching and collecting background information; 4. deliberating about how to proceed; 5. making the choice; and 6. evaluating the results. (For other models proposed in the translation context, see Angelone, 2010; Way, 2014, etc.) On the basis of existing sequential models of problem solving during the translation process, this chapter hypothesises translation problem solving as a process consisting of the following five steps:

1. **Identifying Translation Problems:** Recognising that there is a translation problem to be solved.
2. **Representing Translation Problems:** Understanding the nature of the translation problem and building a problem representation.
3. **Proposing Translation Solutions:** Generating or searching for possible solutions.
4. **Evaluating Translation Solutions:** Evaluating the appropriateness of the proposed translation solutions.
5. **Making Translation Decisions:** Making a decision on a suitable solution for the translation problem.

It is important to emphasise that the proposed translation problem-solving cycle does not imply a strictly sequential order. It only presents an ideal sequence of translation problem solving, which assumes that each step is successfully completed before proceeding to the next step. In the actual translation process, the five steps may occur flexibly, with some being repeated and others being skipped. In order to avoid unnecessary sequential implications, in this chapter, these steps will be referred to as sub-activities (i.e. translation problem-identifying sub-activities, translation problem-representing sub-activities, translation solution-proposing sub-activities, translation solution-evaluating sub-activities, and translation decision-making sub-activities). The proposed cycle differs from previous problem-solving cycles in that the ideal final step does not involve post-choice evaluation, but rather decision making. The consideration behind this difference is that, in the translation process, post-choice evaluation is followed by either changing or retaining the choice, which requires a further decision to be made, and the final decision will be reflected in the translation product. In other words, the translation problem-solving process concludes with a decision-making sub-activity, which links the translation process with the translation product.

Problem-solving ability in translation, therefore, is a procedural ability that orchestrates a combination of knowledge, skills, and attributes to allow each translation problem-solving sub-activity to happen. Consequently, the assessment of this ability requires a process-oriented assessment that pays attention to each translation problem-solving sub-activity. What product-oriented assessment is able to evaluate is merely a limited part of the stage of translation problem-identifying, solution-proposing and decision-making sub-activities. The rest of the iceberg needs to be discovered through the other translation problem-solving sub-activities in the translation process. Moreover, the complexity in the development of this ability also requires a process-oriented assessment, as the progress in the five sub-activities is not necessarily consistent. Product-oriented assessment, therefore, risks neglecting emerging progress as well as mistaking coincidence for progress. It becomes more complicated when the development of relevant knowledge, skills and attributes is considered, as the acquisition of knowledge, and the development of skills and attributes might also be visible only in certain sub-activities.

A LONGITUDINAL STUDY ON PROBLEM SOLVING

To test the relevance of translation problem solving in translation competence, a longitudinal study was conducted with a small group of students enrolled in the MA in Translation and Interpreting Studies (MATIS) programme at the University of Manchester during the academic year 2015/2016. To gather data about the processes of translation problem solving, the longitudinal study used the individual translation task-based interview as its main research instrument, together with two questionnaires and one focus group.

The Translation Task-Based Interview as Main Research Instrument

The main research instrument used for this longitudinal study was the translation task-based interview, which was adapted from the task-based interview proposed by Goldin (1997). While Goldin uses task-based interviews to observe mathematical problem-solving behaviours, the author pointed out the potential of this tool for research on assessment and evaluation. Goldin (1997, pp.61-62) formulates five design principles for a task-based interview including: accessibility, rich representational structure, free problem-solving, explicit criteria, and interaction with the learning environment. Based on these five principles, the translation task-based interview used for this study was designed and constructed as detailed below:

Step 1 (Pre-Translating Interview Section): An English-Chinese translation task was presented to the participant, who was asked to identify translation problems that need to be solved as part of the translation process.

Step 2 (Translating): The participant was asked to complete a translation on a computer, with full access to dictionaries and online resources.

Step 3 (Post-Translating Interview Section): Upon finishing the translation, the participant was interviewed about the translation problem-solving sub-activities carried out during the translation process.

Step 4 (Pre-Revising Interview Section): A prepared peer translation of the same source text was presented to the participant, who was asked to report further identified translation problems as well as to compare and evaluate different solutions.

Step 5 (Revising): The participant was asked to revise the original translation based on the peer translation, with full access to dictionaries and online resources.

Step 6 (Post-Revising Interview Section): After revising the translation, the participant was interviewed about translation problem-solving sub-activities carried out during the revision process.

These steps were followed with each of the six participants who took part in the study, and each interview session was audio-recorded. The translation and revision processes were screen-recorded and observed by the researcher, who also took notes during the observation to anticipate questions for the follow-up interviews (i.e. step 3 and step 6) sections.

The translation task-based interview allowed for the flexible integration of different translation tasks and interview sections at different stages to optimise the observation of translation problem-solving sub-activities. This instrument also enabled the researcher to observe the problem-solving sub-activities carried out by the participants during the actual translation process, and the follow-up interviews (step 3 and step 6) complemented the observation by eliciting more detailed information about both observable and unobservable sub-activities. Each step of the translation task-based interview was carefully designed to gather data on one or more problem-solving sub-activities.

Apart from this main research instrument, another two supporting instruments were designed for the longitudinal study. Two background questionnaires were designed to collect information about the participants' academic background, practical translation experience, career plans, course units of interest and self-evaluation of their current translation competence level. This instrument enabled the researcher to acquire further information regarding possible influencing factors on the participants' problem-solving sub-activities. A focus group was also prepared to gather the participants' feedback on the longitudinal study, opinions about the theoretical perspective, and perspectives on some preliminary findings on

their competence development provided by the researcher. This enabled the researcher to validate some preliminary findings with the participants and gather suggestions for the improvement of future research design. These supporting instruments, together with the main research instrument, allowed the research to obtain substantive data on the translation problem-solving processes carried out by the participants.

The Longitudinal Study: Participants and Implementation

During the academic year 2015/2016, four sessions of translation task-based interviews were conducted respectively with a total of six Chinese students enrolled in the MATIS programme, with an interval of two to three months between interviews. A background questionnaire (see appendix 1) was administered to the participants at the beginning of the study and another follow-up background questionnaire (see appendix 2) was also conducted towards the end of the study. At the end of the study, a focus group (see appendix 3 for a list of questions) was organised with three of the participants attending.

The six participants recruited from the MATIS programme shared the same language combination (i.e. English-Chinese). They had diverse academic backgrounds, and practical translation experience as well as different interests and future plans. These differences resulted in variations in translation problem-solving processes, which enriched the data gathered. Even though the number of participants in this longitudinal study was relatively limited, the four sessions of translation task-based interviews, together with the two questionnaires, and the focus group, provided a comprehensive set of data on problem-solving processes carried out by participants for in-depth qualitative analysis.

The translation tasks which were designed for the four interview sessions consisted of four excerpts of news articles to be translated from English to Chinese. Participants were asked to translate the opening paragraphs (approximately 150 words) of four news articles on relevant topics about China. Each task included a description of the source of the article to be translated along with the corresponding translation brief (including the newspaper/magazine/Web site where the translation would be published). The news articles were selected on the basis of pre-identified translation problems with relevance to the five translation problem-solving sub-activities. In other words, each translation task contained a similar number of pre-identified translation problems that were particularly demanding in each of the translation problem-solving sub-activities. The pre-identification of the translation problems was mainly intended to ensure the comparability of the translation tasks. The translation problems identified by participants during the translation process differed from each other and from those pre-identified by the author.

FINDINGS AND DISCUSSIONS

The data collected about the translation processes followed by the six participants were analysed on the basis of the proposed translation problem-solving sub-activities. The following sections will present and discuss the main findings from an individual analysis of the data, in which the participants' progress (and/or regression) in the sub-activities reveals various aspects of the development of their translation competence during the translation process. In the discussions, the six participants will be referred to as Wang, Tang, Shan, Gong, Wu, and Liu, respectively.

Inconsistent Development Reflected in the Product and the Process

The data analysis revealed that the progress the participants made in different problem-solving sub-activities is not always consistent, leading to differences between the development reflected by the translation process and that the development reflected by the translation product. In many cases, progress is observed in some translation problem-solving sub-activities during the translation process, but such progress is not evident by examining the final translation product. In other instances, the final translation product indicates progress in the problem-solving ability demonstrated by the students, but the data analysis suggests that some sub-activities during the problem-solving process were not performed appropriately, as explained below.

In some cases, the participants were capable of identifying some translation problems, but failed to represent the problems correctly, and consequently, the final translation decisions were inappropriate. In the second of the four interviews conducted with participants, for example, the translation task involved translating the initial paragraphs of a news article from *The Economist* about three influential Chinese liberal economists for *VISTA*, a Chinese news magazine. The title of the news article is ‘Three Wise Men’, which is a concise title alluding to the Bible. The conventions of news titles in the target magazine, however, favour the use of detailed and informative titles to allow the readers to have a good grasp of the content at a glance. Therefore, the participants need to identify and represent this translation problem correctly, in order to propose appropriate translation solutions. In the pre-translating interview, Participant Liu was able to identify the title as a potential translation problem. However, she believed that it was only an intercultural problem with cultural implications and overlooked the fact that this was, rather, a convention-related issue. Consequently, the participant was unable to propose an appropriate translation solution for the news title to comply with the conventions characterising the target magazine.

In some cases, the participants were able to propose translation solutions to the challenges found, but failed to evaluate the solutions effectively, hence still resulting in unsuccessful translation decisions. For example, in the second interview session in which the task involved the translation of a news article on three Chinese economists, the three liberal economists were introduced in the first paragraph as *Mr Market*, *Mr Shareholding* and *the most radical of all, the liberal*, without revealing their full names. With proper context research, the participants should realise that *Mr Market* and *Mr Shareholding* are in fact adapted from the English translation of two economists’ Chinese nicknames (Market Wu and Shareholding Li), and the original Chinese nicknames are therefore the most appropriate translations for the two terms. Participant Gong proposed the two Chinese nicknames at first, but later believed that referring to the economists with their nicknames was not in line with the Chinese discourse conventions. Based on the evaluation, he then decided to adopt a literal translation of the terms used in the news article (i.e. *Mr Market*, *Mr Shareholding*), which, in fact, resulted in an unidiomatic rendering.

In other cases, the participants were able to evaluate translation solutions, yet the lack of a clearly defined macro strategy led to inappropriate translation decisions. For example, in the second session, Participant Shan pointed out that, during the translation process, she constantly tried to recall the stylistic conventions characterising *VISTA* (target magazine) and the style representing the source text authors as well as the stylistic conventions characterising *The Economist* (source magazine). This indicated that in evaluating translation solutions, the participant was capable of considering the translation situation. However, she was not yet able to develop a defined translation strategy, and hesitated about her translation decisions in the post-translating interview, explaining that if more time was available, she would

give some more thought to the style used in the translation to reflect more faithfully the perspective and the style characterising the author of the source text.

Sometimes, when the translation product seemed to indicate that the participants had made progress in their translation problem-solving processes, the problem was not actually identified by the participants. For example, in the third interview session, when translating the sentence *China's notorious claim to nearly the entire South China Sea*, Liu omitted the term *notorious* in her translation into Chinese. This seemed to indicate that she had become aware of the derogatory tone of the source text, which is inappropriate if retained in the target text, as the news translation is intended for Chinese readers. However, in the post-translating interview, Liu revealed that she had simply forgot to translate the word, and pointed out that she should have translated the word to reflect the tone of the source text.

As the examples above demonstrate, the translation product reflects only part of some translation problem-solving sub-activities during the translation process, with many other aspects and other sub-activities only capable of being observed and detected by examining the translation process. The development of translation competence is often more clearly reflected by different translation problem-solving sub-activities during the translation process. The fact that the progress observed from the participants' translation product and translation process is not consistent means that the progress in the translation process is often not immediately shown by the translation product and, therefore, it remains unrecognised if only a product-oriented assessment is conducted. Moreover, when progress is found in the translation product, it does not necessarily indicate progress in the translation problem-solving sub-activities, and may only be a mere coincidence. A process-oriented approach to assessment is, therefore, undoubtedly necessary to gain a full account of the learners' progress in each translation problem-solving sub-activity.

Unbalanced Development in Different Sub-Activities

From a longitudinal view, the progress the participants made in different sub-activities through the four sessions is unbalanced, i.e. it is more significant in some sub-activities than in others. Considering the sub-activities carried out during the four sessions, participant Wang, for example, demonstrated more prominent progress in decision-making and solution-proposing sub-activities. Indeed, her increased confidence in Chinese writing gradually enabled her to propose more contextualised translation solutions with less dependence on external resources in solution-proposing sub-activities, and her gradually increasing knowledge about translation also enabled her to form a personal philosophy of translation which governed her decision-making sub-activities.

In terms of solution-proposing sub-activities, Wang's lack of confidence in her Chinese language abilities played a key role in her reliance on and balance of internal support and external support. Wang completed her undergraduate studies in France, where she did not have the opportunity to practice and improve her Chinese language abilities. In the first session, she showed an over-reliance on online resources when it came to solution-proposing sub-activities. In such cases, the participant constantly consulted Web sites, sometimes being misled by online resources resulting in incorrect translation solutions. When asked why she would copy expressions from Web sites, Wang explained that it was because she was not confident on her Chinese language expression ability. In fact, in the interview sections, she constantly referred to her weaknesses in this language. It is evident that Wang's lack of confidence in linguistic knowledge and skills resulted in the student's over-reliance on external resources. In the second and third sessions, however, the participant began to show less reliance on external resources and limited the consultation of dictionaries and online resources to checks at the beginning of the translation to facilitate

Exploring Process-Oriented Translation Competence Assessment

the understanding of the source text. At this point, Wang was able to propose translation solutions on her own when translating without referring to external resources. Although the participant still felt her weaknesses with the Chinese language were a major issue, the sub-activities seem to indicate that she gradually became more capable of orchestrating her linguistic knowledge and skills to propose satisfactory solutions. The final session evidenced that Wang became more confident in proposing translation solutions based on her Chinese language abilities, as she claimed she felt more capable of expressing herself in Chinese at that point. This regained confidence in Chinese language abilities enabled her to propose translation solutions more efficiently, with a good balance between internal and external resources.

Wang's evolving perception of translation also enabled her to develop a personal philosophy of translation, which assisted her when formulating macro strategies in her decision-making sub-activities. As a trainee translator with little experience prior to the programme, Wang's initial perception of translation was notably influenced by her undergraduate studies in language sciences. Conceiving translation as a process of 'transforming a source text into a target text', Wang believed that a good translation should be fluent and faithful to the source text. Her translation decisions therefore gave prominence to linguistic considerations. In the second session, however, by examining a peer translation, Wang became more aware of register conventions, claiming that the register of the peer translation was more suitable for news articles than the more colloquial tone she had used in her translation. In the revision process, Wang decided that the register of the revised translation should be somewhere between her own 'colloquial translation' and the 'excessively literary' peer translation, and the revisions made by her were, thus, mostly related to the register. In the third session, Wang's perception of translation evolved further, as she began to pay attention to the ideological and professional aspects of the translation situation. Consequently, her macro-translation strategies included strategies to neutralise derogatory expressions and to conform to the conventions of the target Web site. In the final session, Wang's perception of translation became more dynamic and comprehensive. She highlighted that nowadays news translation required the translators 'not to lose themselves in pursuing *xin da ya* (an ancient Chinese translation theory that believes a good translation should be faithful, expressive and elegant)', but to 'be able to domesticate'. In this session, therefore, Wang was no longer constrained by the source text, but was able to translate more flexibly and to make translation decisions more in line with the macro strategy of domestication.

Throughout the four sessions, Wang made steady and significant progress in decision-making and solution-proposing sub-activities. The participant demonstrated a gradual progress in these two sub-activities, with a prominent difference between the first session and the final session. However, the progress in other sub-activities was not as balanced and less noteworthy. Although progress was also observed in these sub-activities, the development was often inconsistent across all the sessions. For example, in solution-evaluating sub-activities, Wang's general development trend was to consider more contextual aspects and make more reasoned evaluation. However, in the final session, Wang's regained confidence in her Chinese language abilities led to a less thorough evaluation of the proposed translation solutions, and a fair number of translation solutions were not evaluated appropriately. Moreover, due to the unbalanced development in these sub-activities, Wang's progress in decision-making and solution-proposing sub-activities was not always detected in the final translation product. Sometimes, the participant failed to identify the translation problem successfully, to evaluate the proposed translation solutions properly, or to follow the macro-translation strategies or principles consistently, which resulted in inappropriate translation decisions in the translation product. In other words, the unbalanced development was only observable by examining the translation process.

For translation teachers, identifying learners' unbalanced development in different translation problem-solving sub-activities during translation processes is conducive to adjusting their teaching focus to develop learners' translation competence more successfully. The identification of the students' strengths and weaknesses will enable teachers to devote more attention to enhance further those sub-activities through which students demonstrate competence development and to encourage progress in other sub-activities which are not as developed. A product-oriented assessment approach, however, is unable to capture the development in different sub-activities with the same level of precision as process-oriented assessment. Wang's case shows it is only through the process that is possible to note she is making consistent and prominent progress in proposing translation solutions and formulating macro-translation strategies, but she still needs to develop further other translation problem-solving sub-activities. A process-oriented assessment approach, therefore, will help highlight the development patterns of the various sub-activities in the translation process.

Individual Differences in Development Paths

Using Wang's case as an example, the preceding section has discussed the different development in translation problem-solving sub-activities carried out by the participants. In fact, an individual-based analysis of the path of development of each participant's translation competence shows that individual differences among the six participants are also prominent. Each participant takes a unique path in his/her translation competence development, which differs in many aspects from other participants' paths.

Individual differences are first manifested in the different translation problem-solving sub-activities, with the participants showing more prominent progress in different sub-activities. Wang made more progress in decision-making and solution-proposing sub-activities, as she became more capable of proposing translation solutions relying on her own language abilities and was more capable of formulating macro strategies based on an expanded perception of translation. Tang's progress was more prominent in problem-identifying sub-activities, as she was able to identify a broader scope of translation problems beyond linguistic and textual problems that were relevant to the translation situation and conventions. Shan made more noteworthy progress in decision-making sub-activities, becoming increasingly capable of formulating and following a macro translation strategy, while Gong experienced competence stagnation, so his problem-solving sub-activities did not show noticeable progress. Wu made relatively notable progress in solution-proposing and solution-evaluating sub-activities, as she became more aware of the need for a balance between internal support and external support. Finally, Liu's progress was more prominent in problem-identifying and decision-making sub-activities, since she became more capable of identifying problems caused by the specific translation situation and became more confident in her translation decisions.

In the same problem-solving sub-activities, development paths also differed. For instance, in solution-proposing sub-activities, both Wang and Wu demonstrated prominent progress, but in opposite directions. Wang's progress in this sub-activity was influenced by the regained confidence in her Chinese language abilities, with which she became more capable of proposing translation solutions on her own instead of overly relying on external resources. Wu's progress in this sub-activity was a result of the waning influence from interpreting. The participant was very enthusiastic about interpreting practice at the beginning of the study, and carried out translation tasks as if she was doing (unprepared) interpreting. She had the habit of not consulting any dictionaries or online resources in her translation process, which often resulted in translation mistakes. However, as her enthusiasm in practising interpreting later

Exploring Process-Oriented Translation Competence Assessment

faded, the participant began to resort to external resources more often during the translation process. Although both Wang and Wu became more capable of striking a better balance between internal support and external support, the starting points and the development tendencies were considerably different.

Individual differences also lay in the different development curves reflected in the translation problem-solving sub-activities carried out by the participants. Some participants demonstrated steady or moderate progress in most of the sub-activities; some made progress in some sub-activities, but showed stagnation or even steps back in other sub-activities, while others showed stagnation in most sub-activities. For example, Wang demonstrated generally steady progress in most sub-activities, although the progress in different sub-activities was unbalanced. Tang made progress in most sub-activities in the first two sessions but, from the third session onwards, she lost confidence in her language abilities, due to negative feedback from the tutor. Stagnation and steps back were thus observed in Tang's solution-proposing and solution-evaluating sub-activities. Gong was a self-taught translation learner with more practical translation experience than other participants at the beginning of the study. With a relatively more established behavioural pattern in translation, he did not demonstrate much progress in any of the sub-activities. Similarly, the other three participants also had very different development curves in their translation problem-solving processes.

The prominent individual differences in the development paths indicate that the development of translation competence is highly individualised. Different translation learners take different or even opposite paths in translation competence development. It is thus necessary for translation teachers to identify and recognise individual differences in competence development so that appropriate feedback and guidance can be given to each learner. This study demonstrates that a product-oriented assessment approach alone does not satisfy this need, since it does not provide information about problem-solving sub-activities during the translation process. A process-oriented assessment approach will be able to assist translation teachers in understanding students' different development paths, based on which learner needs can be more precisely identified and teaching activities can be more targeted to enhance the competence development of each translation learner.

IMPLICATIONS FOR PEDAGOGICAL ASSESSMENT

The analysis of data gathered from the longitudinal study reveals that a process-oriented approach is necessary in the assessment of translation competence and its development. The participants' progress in different problem-solving sub-activities was not always consistent. Such progress is not always apparent by examining the translation product, which reflects only some aspects of progress in some sub-activities. The progress in other aspects and sub-activities that are observable only in the translation process, however, is often more prominent and plays a more crucial role in determining the participants' overall development. Therefore, it is necessary to pay attention to the progress throughout the translation process, to identify not only the overall progress throughout but also an improvement in the product outputs. From a longitudinal view, the progress the participants made in different sub-activities was unbalanced, with some problem-solving sub-activities demonstrating clear progress and others showing less or no progress. It is essential that translation teachers identify strengths and weaknesses in translation processes adopted by learners so that they are able to implement positive reinforcement on emerging or existing development and to perform timely intervention for retrogressive or stagnated development. Moreover, the differences among the participants in their competence development paths

also indicate that competence development is highly individualised, with different learning curves, patterns and development tendencies. Most developmental differences are observable only by examining the translation process. Hence, the empirical findings presented in this chapter highlight the need for a process-oriented approach to competence assessment.

Apart from stressing the necessity to gather developmental evidence from the translation process, the translation problem-solving perspective provides an analytical framework for the evidence gathered. As this chapter highlighted in the introduction, although several process-oriented pedagogical tools have been proposed for the development and assessment of translation competence, the scholarly debates rarely focus on how the evidence gathered can be analysed and interpreted. Further research on the use of such pedagogical tools is therefore paramount. To that effect, the translation problem-solving perspective provides a systematic framework for the analysis of the developmental evidence. Using a translation learning journal, for example, a more systematic analysis can be conducted based on analytical categories adapted from the translation problem-solving sub-activities. More specifically, the contents of learning journals can be categorised according to their relevance to different sub-activities and then analysed within each category or across categories. In a recent article, the author has attempted to apply the analytical framework to analysing a set of learning journals gathered from a specialised translation module at Beijing Foreign Studies University (Cheng and Wu, 2016). This analysis has demonstrated that the analytical framework is effective in mapping translation competence development paths as reflected in the translation process.

The current research also offers new insights for pedagogical tools that can be targeted more effectively at the translation problem-solving process. The translation task-based interview, for instance, provides structural guidance for learning journals. It can be converted and adapted into different forms of structured journal writing, from which translation teachers will be able to gather more developmental evidence for formative assessment. For example, a translation problem report may focus on all sub-activities and require translation learners to structure their journals around the following topics: major translation problems identified in the translation process, nature and structure of identified problems, resources consulted and solutions proposed in the translation process, criteria to evaluate translation solutions, and principles to inform translation decisions. A reflective learning journal may focus on some sub-activities and require translation trainees to reflect upon problem-solving processes with special emphasis on the micro- and macro-strategies adopted in the process. Similarly, a peer-revision journal may focus on some sub-activities and require translation trainees to reflect and write about the evaluation of translation solutions, suggestion of alternative solutions, and macro-translation strategies or principles that guided the final decisions. Such structured translation learning journals will provide key evidence on problem-solving processes, which is conducive to formative process-oriented assessment approaches.

CONCLUSION

This chapter introduces a translation problem-solving perspective to conceptualise translation competence development as reflected in the translation process, which informs the use of process-oriented approaches to translation competence assessment. Findings from a longitudinal study carried out with translation trainees highlight the need for process-oriented assessment approaches and suggest their potential for translation competence development. The conceptual propositions and empirical findings presented in this chapter can contribute to the development of translation pedagogy by providing not

Exploring Process-Oriented Translation Competence Assessment

only theoretical support, but also empirical evidence and practical suggestions for the use and potential of process-oriented assessment approaches in translation teaching and training.

The translation problem-solving perspective also bridges translation competence research and relevant research within and beyond the translation context. In translation studies, numerous findings from translation process research, especially those related to translation expertise and learner-professional comparisons, can be interpreted in problem-solving terms and incorporated into the proposed conceptualisation to enrich the current understanding of translation competence and its development. The translation problem-solving perspective also encourages more interdisciplinary research to be performed with relevant disciplines, such as expertise studies and development studies. With a focus on problem-solving abilities, these disciplines have generated many findings that will deepen the understanding of translation competence and its development as part of the learning process. In existing literature, there is an emerging tendency to introduce relevant findings from these disciplines into translation competence research (e.g. Angelone, 2010; Göpferich & Jääskeläinen, 2009; Hansen, 2010; Prieto Ramos, 2011; Shreve, 2002; Way, 2014), but there is a pressing need for a systematic framework for the integration of these findings. The translation problem-solving perspective enables more relevant findings to be incorporated into translation competence research and hence enrich the understanding of translation competence and its development.

The longitudinal study presented in this chapter also offers inspiration for future empirical research. Based on the proposed translation task-based interview, similar research instruments can be devised to observe one or more sub-activities in the participants' translation and revision processes. For instance, a revision task-based interview can be designed to gather data on self-revision or peer-revision processes. Similar longitudinal studies can be conducted using control groups that are informed of the translation problem-solving perspective, to reveal the influence of explicit knowledge about translation problem solving on the development of the translation problem-solving ability. Alternatively, further longitudinal studies can be conducted with participants with different levels of translation competence in translation programmes of different length, so that the development patterns at different stages and the development paths over a longer period of learning can be identified and understood.

REFERENCES

- Angelone, E. (2010). Uncertainty, uncertainty management and metacognitive problem solving in the translation task. In G. M. Shreve & E. Angelone (Eds.), *Translation and cognition* (pp. 17–40). Amsterdam: John Benjamins. doi:10.1075/ata.xv.03ang
- Angelone, E. (2016). A process-oriented approach for documenting and gauging intercultural competence in translation. *The Interpreter and Translator Trainer*, 10(3), 304–317. doi:10.1080/1750399X.2016.1236560
- Bell, R. T. (1991). *Translation and translating: Theory and practice*. New York: Longman London.
- Bell, R. T. (1998). Psycholinguistic/cognitive approaches. In M. Baker (Ed.), *Routledge encyclopedia of translation studies* (pp. 185–190). London: Routledge.
- Biel, Ł. (2011). Professional realism in the legal translation classroom: Translation competence and translator competence. *Meta*, 56(1), 162–178. doi:10.7202/1003515ar

Campbell, S. (1991). Towards a model of translation competence. *Meta*, 36(2-3), 329–343. doi:10.7202/002190ar

Campbell, S. (1998). *Translation into the second language*. London: Longman.

Cheng, S., & Wu, Q. (2016). Mapping the development of translation competence in learning journals: A dynamic dimension and analytical framework inspired by problem-solving studies. *Chinese Translators Journal*, 1, 51–57.

Darwish, A. (2008). *Optimality in translation*. Victoria: Writescope Pty Ltd.

Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. New York: Free Press.

European Master's in Translation (EMT). (2009). *Competences for professional translators, experts in multilingual and multimedia communication*. Retrieved January 5, 2018, from https://ec.europa.eu/info/sites/info/files/emt_competences_translators_en.pdf

Fox, O. (2000). The use of translation diaries in a process-oriented translation teaching methodology. In C. Schäffner & B. Adab (Eds.), *Developing translation competence* (pp. 115–130). Amsterdam: John Benjamins. doi:10.1075/btl.38.12fox

Gaddis Rose, M. (1979). Translation as problem-solving. *Translation Review*, 3(1), 20–21. doi:10.1080/07374836.1979.10523556

Gile, D. (2004). Integrated problem and decision reporting as a translator training tool. *The Journal of Specialised Translation*, 2, 2–20.

Gile, D. (2009). *Basic concepts and models for interpreter and translator training* (Revised ed.). Amsterdam: John Benjamins. doi:10.1075/btl.8

Goldin, G. A. (1997). Observing mathematical problem solving through task-based interviews. *Qualitative Research Methods in Mathematics Education*, 9, 40–62.

Göpferich, S. (2009). Towards a model of translation competence and its acquisition: The longitudinal study TransComp. In S. Göpferich, A. L. Jakobsen, & I. M. Mees (Eds.), *Behind the mind: Methods, models and results in translation process research* (pp. 11–37). Copenhagen: Samfundslitteratur Press.

Göpferich, S. (2013). Translation competence: Explaining development and stagnation from a dynamic systems perspective. *Target*, 25(1), 61–76. doi:10.1075/target.25.1.06goe

Göpferich, S., & Jääskeläinen, R. (2009). Process research into the development of translation competence: Where are we, and where do we need to go? *Across Languages and Cultures*, 10(2), 169–191. doi:10.1556/Acr.10.2009.2.1

Hansen, G. (2010). Integrative description of translation processes. In G. M. Shreve & E. Angelone (Eds.), *Translation and cognition* (pp. 189–211). Amsterdam: John Benjamins. doi:10.1075/ata.xv.11han

Hayes, J. R. (1989). *The complete problem solver* (2nd ed.). Lawrence Erlbaum Associates, Inc.

House, J. (1997). *Translation quality assessment: A model revisited*. Tübingen: Gunter Narr Verlag.

Exploring Process-Oriented Translation Competence Assessment

Huertas Barros, E., & Vine, J. (2018a). Current trends on MA translation courses in the UK: Changing assessment practices on core translation modules. *The Interpreter and Translator Trainer*, 12(1). Retrieved from <http://www.tandfonline.com/doi/full/10.1080/1750399X.2017.1400365>

Huertas Barros, E., & Vine, J. (2018b). Constructing standards in communities: Tutors' and students' perceptions of assessment practices on an MA translation course. In E. Huertas Barros, S. Vandepitte & E. Iglesias Fernández (Eds.), *Quality assurance and assessment practices in translation and interpreting: Advances in linguistics and communication studies (ALCS)*. Hershey, PA: IGI Global.

Hurtado Albir, A. (2017). Translation and translation competence. In A. Hurtado Albir (Ed.), *Researching translation competence by PACTE Group* (pp. 3–34). Amsterdam: John Benjamins. doi:10.1075/btl.127.01hur

Kelly, D. (2007). Translator competence contextualized. Translator training in the framework of higher education reform: In search of alignment in curricular design. In D. Kenny & K. Ryou (Eds.), *Across boundaries: International perspectives on translation studies* (pp. 128–142). Cambridge, UK: Cambridge Scholars Publishing.

Kim, M. (2009). Meaning-oriented assessment of translations. In C. V. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies* (pp. 123–157). Amsterdam: John Benjamins. doi:10.1075/ata.xiv.08kim

Kim, M. (2013). Research on translator and interpreter education. In C. Millán & F. Bartrina (Eds.), *The Routledge handbook of translation studies* (pp. 102–116). London: Routledge.

Kiraly, D. (1995). *Pathways to translation: Pedagogy and process*. Kent, OH: The Kent State University Press.

Kiraly, D. (2000). *A social constructivist approach to translator education: Empowerment from theory to practice*. Manchester, UK: St Jerome Publishing.

Kiraly, D. (2013). Towards a view of translator competence as an emergent phenomenon: Thinking outside the box (es) in translator education. In D. Kiraly, S. Hansen-Schirra, & K. Maksymski (Eds.), *New prospects and perspectives for educating language mediators* (pp. 197–224). Tübingen: Gunter Narr Verlag.

Kiraly, D., & Piotrowska, M. (2014). *Towards an emergent curriculum development model for the European Graduate Placement Scheme*. Paper presented at the Future of Education, Florence, Italy.

Krings, H. P. (1986). Translation problems and translation strategies of advanced German learners of French (L2). In J. House & S. Blum-Kulka (Eds.), *Interlingual and intercultural communication* (pp. 263–276). Tübingen: Gunter Narr Verlag.

Kusmaul, P. (1995). *Training the translator*. Amsterdam: John Benjamins. doi:10.1075/btl.10

Levý, J. (2000). Translation as a decision process. In V. Lawrence (Ed.), *The translation studies reader* (pp. 148–159). London: Routledge.

Martínez Melis, N., & Hurtado Albir, A. (2001). Assessment in translation studies: Research needs. *Meta*, 46(2), 272–287. doi:10.7202/003624ar

- Massey, G. (2005). Process-oriented translator training and the challenge for E-learning. *Meta*, 50(2), 626–633. doi:10.7202/011006ar
- National Accreditation Authority for Translators and Interpreters (NAATI). (2015). *NAATI translator certification: Knowledge, skills and attributes*. Retrieved June 3, 2017, from https://www.naati.com.au/media/1059/translator_ksa_paper.pdf
- Neubert, A. (1994). Competence in translation: A complex skill, how to study and how to teach it. In M. Snell-Hornby, F. Pöchhacker, & K. Kaindl (Eds.), *Translation studies: An interdisciplinary* (pp. 411–420). Amsterdam: John Benjamins. doi:10.1075/btl.2.48neu
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.
- Nord, C. (1992). Text analysis in translator training. In C. Dollerup & A. Loddegaard (Eds.), *Teaching translation and interpreting: Training, talent, and experience* (pp. 39–48). Amsterdam: John Benjamins. doi:10.1075/z.56.08nor
- Orlando, M. (2012). Training of professional translators in Australia: Process-oriented and product-oriented evaluation approaches. In S. Hubscher-Davidson & M. Borodo (Eds.), *Global trends in translator and interpreter training* (pp. 197–216). London: Continuum.
- PACTE. (2003). Building a translation competence model. In F. Alves (Ed.), *Triangulating translation* (pp. 43–66). Amsterdam: John Benjamins.
- PACTE. (2005). Investigating translation competence: Conceptual and methodological issues. *Meta*, 50(2), 609–619. doi:10.7202/011004ar
- PACTE. (2017). The second stage of PACTE group’s research: Experimental research into the acquisition of translation competence. In A. Hurtado Albir (Ed.), *Researching Translation Competence by PACTE Group* (pp. 303–308). Amsterdam: John Benjamins.
- Pretz, J. E., Naples, A. J., & Sternberg, R. J. (2003). Recognizing, defining, and representing problems. In J. E. Davidson & R. J. Sternberg (Eds.), *The psychology of problem solving*. New York: Cambridge University Press. doi:10.1017/CBO9780511615771.002
- Prieto Ramos, F. (2011). Developing legal translation competence: An integrative process-oriented approach. *Comparative Legilinguistics - International Journal for Legal Communication*, 5, 7-21.
- Prieto Ramos, F. (2014). Parameters for problem-solving in legal translation: Implications for legal lexicography and institutional terminology management. In L. Cheng, K. K. Sin, & A. Wagner (Eds.), *The Ashgate handbook of legal translation* (pp. 121–134). London: Routledge.
- Pym, A. (1992). Translation error analysis and the interface with language teaching. In C. Dollerup & A. Loddegaard (Eds.), *Teaching translation and interpreting* (pp. 279–288). Amsterdam: John Benjamins. doi:10.1075/z.56.42pym
- Pym, A. (2003). Redefining translation competence in an electronic age. In defence of a minimalist approach. *Meta*, 48(4), 481–497. doi:10.7202/008533ar

Exploring Process-Oriented Translation Competence Assessment

Pym, A. (2011). Training translators. In K. Malmkjær & K. Windle (Eds.), *The Oxford handbook of translation studies* (pp. 475–489). New York: Oxford University Press.

Robertson, S. I. (2001). *Problem solving*. Philadelphia: Psychology Press. doi:10.4324/9780203457955

Schäffner, C., & Adab, B. (2000). Developing translation competence: Introduction. In C. Schäffner & B. Adab (Eds.), *Developing translation competence* (Vol. 38, pp. vii–xvi). Amsterdam: John Benjamins. doi:10.1075/btl.38.01sch

Shei, C. C.-C. (2005). Translation commentary: A happy medium between translation curriculum and EAP. *System*, 33(2), 309–325. doi:10.1016/j.system.2005.01.004

Shreve, G. M. (2002). Knowing translation: Cognitive and experiential aspects of translation expertise from the perspective of expertise studies. In A. Riccardi (Ed.), *Translation studies: Perspectives on an emerging discipline* (pp. 150–171). New York: Cambridge University Press.

Sirén, S., & Hakkarainen, K. (2002). Expertise in translation. *Across Languages and Cultures*, 3(1), 71–82. doi:10.1556/Acr.3.2002.1.5

Tirkkonen-Condit, S. (2000). Uncertainty in translation processes. In S. Tirkkonen-Condit & R. Jääskeläinen (Eds.), *Tapping and mapping the processes of translation and interpreting: Outlooks on empirical research* (pp. 123–142). Amsterdam: John Benjamins. doi:10.1075/btl.37.13tir

Way, C. (2014). Structuring a legal translation course: A framework for decision making in legal translator training. In L. Cheng, K. K. Sin, & A. Wagner (Eds.), *The Ashgate handbook of legal translation* (pp. 135–154). London: Routledge.

Wilss, W. (1976). Perspectives and limitations of a didactic framework for the teaching of translation. In R. W. Brislin (Ed.), *Translation: Applications and research* (pp. 117–137). New York: Gardner Press.

Wilss, W. (1996). *Knowledge and skills in translator behavior*. Amsterdam: John Benjamins. doi:10.1075/btl.15

Zhong, Y. (2005). Plan-based translation assessment: An alternative to the standard-based cut-the-feet-to-fit-the-shoes style of assessment. *Meta*, 50(4).

ENDNOTE

- ¹ The terms ‘competence’ and ‘competency’ have been used in this chapter to retain the respective terms used by different scholars in the existing literature.

APPENDIX 1

First Background Questionnaire

Thank you for agreeing to fill in this questionnaire. This questionnaire is designed to find out a few things about your translation learning and practice. It is part of a PhD project on translation competence development. Please note that the personal information collected in this questionnaire will only be used for the purpose of building your portfolio during the research. Anonymity will be guaranteed.

How long will it take?

We would appreciate it if you could spare five to ten minutes of your time to fill in this questionnaire.

What about confidentiality?

The information collected will only be used for academic purposes. The information will not be passed on to any third party.

What do I have to do?

We ask you to please answer the questions truthfully. There are no right or wrong answers.

Part I: Your Previous Experience

1. Do you have a previous degree in translation and interpreting (T&I)?
 Yes. I hold a BA/ MA degree in T&I/translation studies.
 No. I hold a BA/ MA degree in _____ (please specify your major)
2. Have you received any practical T&I training before starting this programme?
 Yes, from university courses training institutes
 No.
 Other.
3. Have you learned any translation theories before starting this programme?
 Yes, please list up to three theoretical concepts in translation studies that you are familiar with

 No.
4. Have you had any freelance/in-house T&I experience before starting this programme?
 Yes, I have translated approximately _____ (e.g. 10k) words in fields including _____ (e.g. literary, commercial, etc.) as a freelance/in-house (please circle) translator.
 No.
5. Have you taken any T&I qualification exams in China (e.g. CATTI, NAETI)?
 Yes. Please specify _____
 No.

Exploring Process-Oriented Translation Competence Assessment

6. From 1 to 5, how would you grade your own translation competence, where 1 is no competence at all, while 5 is extremely competent? (Please circle the number below)
- 1 2 3 4 5

Part II: Your Interests and Plans

7. Which of the following course modules do you have an interest in?

Please find introductions to these course modules from the programme handbook. Note that this question asks about what course modules you are interested in taking, not what course modules you are planning to take.

- Research Methods in T&I Studies I
 - Translation and Interpreting Studies I
 - Translating for International Organizations
 - Commercial Translation
 - Translation Technologies
 - Consecutive Interpreting
 - Research Methods in T&I Studies II
 - Cross-Cultural Pragmatics
 - Literary Translation: Contexts and Challenges
 - Audiovisual Translation I
 - Scientific and Technical Translation
 - Translation and Media Culture
 - Translation Project Management and Professional Ethics
 - Public Service Interpreting
8. Which of the following dissertation type might you choose to do?
- Research dissertation
 - Practical dissertation (translation consecutive interpreting public service interpreting subtitling)
 - Not sure yet
9. Are you planning to do any freelance T&I work during the programme?
- Yes, I am planning to work in these fields: _____
 - No.
 - Not sure.
10. Are you planning to work as a professional translator/interpreter after graduation?
- Yes. (in-house translator/interpreter freelance translator/interpreter)
 - No. (PhD/further study T&I-related profession other profession)
 - Not sure.

Part III: Your Personal Information

Name _____

Email _____

Thank you for completing this questionnaire!

APPENDIX 2

Second Background Questionnaire

Thank you for agreeing to fill in this questionnaire. This questionnaire is designed to find out a few things about your translation learning and practice. It is part of a PhD project on translation competence development. Please note that the personal information collected in this questionnaire will only be used for the purpose of building your profile during the research. Anonymity will be guaranteed.

How long will it take?

We would appreciate it if you could spare five minutes of your time to fill in this questionnaire.

What about confidentiality?

The information collected will only be used for academic purposes. The information will not be passed on to any third party.

What do I have to do?

We ask you to please answer the questions truthfully. There are no right or wrong answers.

Part I: Your MATIS Programme Experience

1. Which of the following course modules have you taken?
 - Research Methods in T&I Studies I
 - Translation and Interpreting Studies I
 - Translating for International Organisations
 - Commercial Translation
 - Translation Technologies
 - Consecutive Interpreting
 - Research Methods in T&I Studies II
 - Cross-Cultural Pragmatics
 - Literary Translation: Contexts and Challenges

Exploring Process-Oriented Translation Competence Assessment

- Audiovisual Translation I
 - Scientific and Technical Translation
 - Translation and Media Culture
 - Translation Project Management and Professional Ethics
 - Public Service Interpreting
2. Which of the following dissertation types have you chosen? Please provide some key words of your dissertation (topic, theory, *etc.*).
- Research dissertation _____
 - Practical dissertation (translation consecutive interpreting public service interpreting Subtitling) _____
3. From 1 to 5, how would you grade your own translation competence after completing this programme, where 1 is no competence at all, while 5 is extremely competent? (Please circle the number below)
- 1 2 3 4 5

Part II: Your Extracurricular Translation Experience

4. Have you had any freelance/in-house translation experience during this programme?
- Yes, I have translated approximately ____ (*e.g.* 10k) words in fields including _____ (*e.g.* literary, commercial, *etc.*) as a freelance/in-house (please circle) translator.
- No.
5. Have you had any freelance/in-house interpreting experience during this programme?
- Yes, I have interpreted approximately ____ hours in fields including _____ (*e.g.* literary, commercial, *etc.*) as a freelance/in-house (please circle) interpreter.
- No.
6. Have you taken any T&I qualification exams during this programme?
- Yes. Please specify _____
- No.
7. Are you planning to work as a professional translator or interpreter after graduation?
- Yes. (in-house translator in-house interpreter freelance translator freelance interpreter)
- No. (PhD/further study. Please specify _____
- T&I-related profession. Please specify _____
- other profession. Please specify _____)
- Not sure.

Part III: Your Personal Information

Name _____

Email _____

Thank you for completing this questionnaire!

APPENDIX 3

List of Focus Group Questions

[These questions were originally presented to the participants in Chinese and translated here by the author.]

Part I

Theme 1: On the Translation Task-Based Interview

First of all, I would like to thank all of you for attending all the four translation task-based interview sessions during the past academic year. Now, please recall your experience in the four interview sessions and share your thoughts on the following topics:

- How did you feel about the interview sessions?
- In the interview sessions, did you translate like you would normally translate in your previous translation activities? If not, what were the major differences?
- Did you reflect upon your translation performance in the interview sessions after each session?
- Have you benefitted from attending the interview sessions? If yes, what are the benefits?
- Do you have any suggestions for improving the design of the interview sessions?

Theme 2: On the Access to Others' Translations

In each interview session, you had access to a peer translation. In your previous translation activities, you might also have found existing translations online. Please share your thoughts on the following topics:

- In your translation process, do you wish to have access to others' translations? Why, or why not?
- When do you wish to have access to others' translations, before your translation process starts, during your translation process or after your translation process ends?

Theme 3: On the Access to Others' Translation Processes

Please share your thoughts on the following topics:

- When reading others' translations, are you curious about the reason for some translation choices?
- Do you wish to have access to others' translation processes, especially the processes of identifying and solving translation problems (e.g. what online resources are consulted, what adjustments are made)?
- Which do you wish to have more access to, others' final translation products, or their translation processes? Why?

Theme 4: On Translation Commentaries/Journals/Reports

Please share your thoughts on the following topics:

- In the specialist translation course units, do your instructors/tutors' evaluations of your translation products meet your own expectation?

Exploring Process-Oriented Translation Competence Assessment

- In the specialist translation course units, do you think it is necessary to submit a translation journal (e.g. to point out translation problems, to explain translation choices) along with the translation product? Why?

Theme 5: On the Translation Problem-Solving Process

Some researchers have proposed that there are five sub-activities in the translation problem-solving process: identifying translation problems, representing translation problems, proposing translation solutions, evaluating translation solutions, and making translation decisions. Please share your thoughts on the following topic:

- Do you think the five sub-activities have covered your activities in your translation processes? If not, what are the other activities that have not been included?

Part II

Theme 6: On Some Preliminary Findings

Please review your general translation activities during the past academic year, and discuss to what extent you agree or disagree with the following statements.

- I am more capable of identifying a wider range of translation problems (e.g. translation situation-related problems, contextual knowledge-related problems) in my translation activities now than one year ago.
- When I encounter new translation problems, I am more capable of recalling similar problems identified and solved before to solve the current problems more efficiently now than one year ago.
- When I encounter new translation problems, I am more capable of analysing (the type, nature or cause of) identified translation problems and retrieving relevant solving strategies now than one year ago.
- I have a more critical attitude towards the credibility of online resources and tend to read information from more reliable sources now than one year ago.
- I am less dependent on translations provided by dictionaries or online resources and more flexible in proposing my own translation solutions now than one year ago.
- I take more consideration of the specific translation situation (e.g. the commissioner, the target readership) when I evaluate alternative translation solutions now than one year ago.
- I follow more defined translation principles and translation strategies in my translation activities now than one year ago.
- I like to reflect upon the problems I have identified and solved in my translation processes and summarise the solving strategies.

Chapter 10

Cognitive Effort and Efficiency in Translation Revision

Moritz Schaeffer

Johannes Gutenberg-Universität Mainz, Germany

Anke Tardel

Johannes Gutenberg-Universität Mainz, Germany

Sascha Hofmann

Johannes Gutenberg-Universität Mainz, Germany

Silvia Hansen-Schirra

Johannes Gutenberg-Universität Mainz, Germany

ABSTRACT

Empirical studies of revision are often based on either think aloud protocols, interviews, or observational methods. Eye tracking and keylogging methods are rarely applied to the study of revision behavior. The authors employ established methods from translation process research (TPR) to study the eye movement and typing behavior during self-revision (i.e., the phase in the translation process that follows a first complete draft). The authors measure the effect of behavior during the drafting phase on the relative revision duration. Relative revision duration is the time translators spend revising the first complete draft of the source text. They find that the most efficient process involves a large degree of concurrent reading and writing and few deletions during the drafting phase. The efficiency gains in terms of relative revision duration achieved by avoiding discontinuous typing, by making a larger number of deletions, pausing for longer amounts of time, and engaging in less concurrent reading and writing are outweighed by the gains in total task time by doing the exact opposite.

INTRODUCTION

Revision is today undoubtedly a major quality management task within the translation process (cf. Gouadec, 2007; Mossop, 2007; Drugan, 2013; ISO 17100, 2015). Producing a product without checking its quality would, according to the ISO 9001 requirements of measuring process results (ISO 9001,

DOI: 10.4018/978-1-5225-5225-3.ch010

2015), simply not be allowed. Even if we disregard standards and only take a look at a typical modern translation production, this would be nothing other than negligent. The field of research that evolves out of this standard-driven necessity to revise translations is the investigation of the translation revision process. For a more detailed process description, the identification of the different process parts and the definition of the most efficient revision mode are very important. In this paper we will present a study that investigates self-revision within the translation revision process. The main research question addresses the revision behavior and its efficiency during self-revision performed by the translator.

Almost exactly ten years ago in an article summing up the status of empirical studies in revision, Mossop (2007, p. 18) expected eye tracking to be the future technology to allow translation revision research to combine and triangulate data from different sources for a better investigation of the revision process. One key proposal was to correlate keystroke records, at the time a very common method in research, with the recordings from eye movements of the revision actions that are performed (or not) on a translated text. While his 2007 publication was limited to the research question why revisers would overlook certain errors, he foresaw the possibilities of correlating eye movements and keystrokes in combination with an empirical set of methods to investigate the behaviour during the different phases of the translation process. Although there are a number of studies investigating revision and although several behavioural aspects were investigated empirically (e.g. Mossop, 2014; Künzli, 2005, 2006, 2009, 2014; Robert, 2008, 2014; Robert, Remael, & Ureel, 2016), eye tracking and key logging have only been used in some smaller scale studies in revision research (Englund Dimitrova, 2005; Robert, 2012)¹.

In Translation Process Research (TPR), the written translation process as a whole is typically divided into an orientation, a drafting and a final revision phase (Jakobsen, 2002). We define, for the current purpose, the drafting phase as starting with the first keystroke. The drafting phase ends with the first time that the last source word was translated. All the reading that occurs prior to the drafting phase is considered orientation and all the keystrokes and reading that occur after the drafting phase are considered end revision. Indeed we distinguish two types of revision: revision during the drafting phase (online revision) and end revision, which are corrections that take place during the final revision phase. This common understanding of revision in the TPR community seems to be similar to the concept of self-revision or checking by the translator in revision research.

EMPIRICAL STUDIES OF REVISION

There is a limited number of empirical studies which investigate the translation process as opposed to the product. The data elicitation methods in these studies are varied. They include interview studies (e.g. Shih, 2006), Think Aloud Protocols (TAPs) (e.g. Shih, 2013), screen recording and key logging and/or eye tracking. Given that it has been shown that TAPs have a distorting effect on the process of translation itself (Jakobsen, 2003) we will, for the current purpose, ignore studies which use TAPs as data elicitation method. Shreve, Angelone, and Lacruz (2014) study to what extent viewing a screen recording of someone else's translation process can help spotting errors in the product and find that it is indeed more helpful than other methods. Ipsen and Dam (2016) investigate which revision procedure is the most successful (procedures were, e.g., whether the source or the target were read first). Success here is defined as error detection which is, to a certain extent, independent of how the error is corrected. Results suggest that the more successful procedures are the ones which involve reading the target text (TT) first. However, given that the study by Ipsen and Dam (2016) used screen recording and post-task

interviews only, these results have to be interpreted cautiously. It can be argued that it is impossible to know what was read on the basis of screen recordings and post-task interviews since they may not be the most reliable source of information regarding processes which occur very fast, possibly without participants being aware of them and thus easily forgotten. Malkiel (2009) uses key logging in order to identify corrections in the flow of text production. However, the durations of these corrections are not analysed. Malkiel analyses the nature of the error corrections in regards to linguistic aspects. But while it is possible to hypothesise about the probable cause of the error on the basis of an error typology, it is difficult to conclude anything regarding the associated cognitive effort involved in the error production or recognition and following revision if durations are not analysed. Alves and Vale (2011) proceed in a similar manner, i.e., they use key logging in order to reconstruct and analyse the corrections which occur during text production. Alves and Vale coin the terms micro and macro units. A macro unit consists of at least two micro units in which the same words are revised / corrected. Micro units are constituted by coherent typing activity separated by pauses of at least five seconds. In a vein very similar to the study by Malkiel (2009), Alves and Vale (2011) offer a detailed analysis of the TT and its interim solutions, but it is difficult to conclude anything regarding cognitive effort on this basis, given that the durations of these micro units are not analysed. However, Alves and Vale show that the two extreme types of revision behaviour – either many or almost all revisions taking place during the drafting phase or revisions only or predominantly taking place during the final revision phase are rare in their sample. It is unclear which process is actually the most efficient process. The study by Carl, Schaeffer, and Bangalore (2010) is based mainly on key logging but also on eye tracking data from 24 participants (12 students and 12 professionals). The authors find that there is a tendency for students to spend more time on initial orientation and less time on the final revision phase whereas for professionals, the opposite seems to be the case. However, these tendencies were not tested statistically. On the basis of eye movement and key logging data from 24 participants (12 students and 12 professionals), Dragsted and Carl (2013) find that some participants show more regressions to already translated text than others during the drafting stage. This can be seen as online revision, i.e., translators re-read translated text during the drafting phase in order to make sure that the typed text does not contain errors.

RESEARCH QUESTION

Given the importance of the final revision phase (ISO 17100, 2015) in general and in regards to the quality of the final TT, it seems only natural to study the aspects which have an effect on this phase of the translation process. The translation process can be understood as a process with bottlenecks, i.e., it can be understood as a process which generally runs smoothly and at certain points, the smooth running is more or less severely restricted. How the translation process as a whole is organised can have an effect on its efficiency. It may well be, for example, that it is more efficient to produce a first quick draft leaving all unknown words in the interim draft and only research them after the interim draft is completed. An alternative organisation of the process would be to research unknown words as and when they are read. Unknown words and the need to carry out relevant research in external sources could be seen as bottlenecks and how the bottlenecks are dealt with in the process can also have an effect on the overall efficiency. We define efficiency, for the present purpose, as the process which is the fastest, as measured by task completion time. The data we use (see below) consists of translations from English into 6 different target languages.

Cognitive Effort and Efficiency in Translation Revision

Finding the most efficient process has a range of possible applications such as optimising translator training or the production process of established professional translators. If it is found that research in external sources is more efficient after a first rough draft rather than during the drafting phase, this could be taught to students during training or to professionals who might not be aware of this potential gain in efficiency. Existing empirical studies do not address the question of whether it is more efficient to spend much time and effort on online revision and little time and effort on final revision or vice versa.

Our main aim is to find factors which can predict revision behaviour. Common sense tells us that making mistakes during the drafting stage is inevitable to a certain extent. However, how those mistakes are addressed may have an effect on the overall efficiency of the process as a whole, as measured by the task completion time. In other words, the aim of this chapter is to find out how to produce a translation involving a final revision phase in the quickest possible way.

In general, we hypothesise that the two tasks of online revision during drafting and revision in a final revision phase are related in a way that translators either revise more during drafting or translate continuously during drafting and spend more time and effort on the final revision phase of the final TT once a first draft is completed which is then relatively longer. In addition, we hypothesise that online revision has an impact on the overall duration of the translation process and thus inhibits efficiency. We will test these hypotheses on the basis of a subset of the TPR-DB (Carl et al., 2016).

The findings of this study are relevant for the practice of the profession, given that they could form the basis for guidelines to professionals and students in terms of whether it is more conducive for the efficiency of the translation process to revise more during the drafting phase or to spend proportionally more time on the final revision phase. We believe that it is time to extend the well-established research methods and procedures from TPR to revision to get further insights into another phase of the core translation process, i.e. the final revision phase. Extending the knowledge base that mainly consists of surveys, interviews and observational studies (e.g. Shih, 2006; Mossop, 2007; Shih, 2013; Künzli, 2014) seems a particularly promising avenue.

DATA

We used the MultiLing² dataset from the TPR-database (Carl et al., 2016) and only included the translation from scratch sessions. Table 1 shows a summary of descriptive statistics for all translation tasks per language.

There are 348 sessions (332 after the exclusion of outliers) of 164 participants (117 students and 47 professionals) in total. Each participant translated three of the six English source texts (ST) which are ~150 words in length (between 5 and 11 sentences) and of a relatively general nature: 4 newspaper articles and 2 encyclopedia entries. The brief for most studies in the database was very general, i.e., participants were told to translate to the best of their knowledge at the highest possible quality. Layout did not play a role given that Translog does not mirror the options a standard Word document offers.

Data Analysis

For the data analysis we used *Linear Mixed Effect Models* (LME) fitted with REML and tested a range of predictors from the drafting and final revision phases that indicated reading and text production. In this chapter we present two final models which are the most parsimonious ones and only include predictors

Table 1. Number and percentages of participants and translation sessions and distribution of status per language

Target Language	Participants (Student/Prof)	Sessions (All)	Sessions (Students)	Sessions (Professional Translators)	Sessions of Participants With <2 Years' Experience
Spanish	31 (26/5)	60 (17.2%)	50 (83.3%)	10 (16.7%)	36 (60.0%)
Japanese	36 (17/19)	66 (19.0%)	33 (50.0%)	33 (50.0%)	26 (39.4%)
Danish	24 (12/12)	68 (19.5%)	33 (48.5%)	35 (51.5%)	33 (48.5%)
Hindi	20 (20/-)	36 (10.3%)	36 (100%)	-	2 (5.6%)
Chinese	29 (29/-)	72 (20.7%)	72 (100%)	-	70 (97.2%)
German	24 (13/11)	46 (13.2%)	26 (56.5%)	20 (43.5%)	26 (56.5%)
Total	164 (117/47)	348 (100%)	250 (71.8%)	98 (28.2%)	193 (55.5%)

that had a significant effect. For the LME fitting we used *languageR* (R Development Core Team, 2014) and the packages *lme4* (Bates et al., 2014), *lmerTest* (Kuznetsova, Christensen, & Brockhoff, 2014) to test for significance, and *effects* (Fox et al., 2016) to visualize the effects of the models. Prior to fitting, we looked at the non-parametric curves for each predictor. If the curve indicated a positive or negative tendency, the variables were tested in the LME. Predictors that showed a significant effect in the model, were also checked for normal distribution with the *Shapiro Wilk Test*. As is often the case with natural data, they were not normally distributed. We decided to exclude outliers for all predictors which were more than 3.0 standard deviations above or below the overall mean, i.e., the mean across all translation sessions, for that variable. In total, we excluded 4.6% data points. We then checked again for normal distribution and found that log transformation did not lead to normal distribution either. In the end, we refitted the model with the dataset excluding outliers. A descriptive summary statistic after outliers were excluded is given in Table 2 for overall task duration and final revision phase duration.

The final model fit for both models was tested against a null model. Using the package *MuMIn* (Bartoń, 2009) which enables r^2 calculation for the model with fixed effects only (marginal R^2) and including random effects (conditional R^2), we report to what extent the model explains the variance in

Table 2. Descriptive summary statistics after exclusion of outliers for mean task duration, revision phase duration (in minutes and %) and number of sessions per language (standard deviations and max values in brackets)

Language	Mean Task Duration (in Minutes)	Mean End-Revision Duration (in Minutes)	Mean % Revision (in % of Task Duration)	Sessions
Spanish	9.9 (SD 3.1, Max 19.8)	2.0 (SD 1.4, Max 6.3)	20.4 (SD 11.4, Max 55.9)	59
Japanese	18.3 (SD 6.7, Max 36.1)	3.0 (SD 3.1, Max 15.1)	16.0 (SD 13.0, Max 54.9)	66
Danish	7.0 (SD 2.1, Max 14)	1.0 (SD 0.9, Max 4.0)	14.0 (SD 10.3, Max 42.4)	65
Hindi	16.9 (SD 6.5, Max 34.9)	2.7 (SD 2.5, Max 10.4)	15.1 (SD 11.6, Max 35.5)	27
Chinese	14.1 (SD 4.1, Max 25.3)	1.8 (SD 1.7, Max 6.0)	11.7 (SD 10.5, Max 37.5)	72
German	16.5 (SD 5.1, Max 26.9)	3.1 (SD 2.1, Max 8.4)	18.6 (SD 10.8, Max 37.5)	43
Total	13.3 (SD 6.2, Max 36.1)	2.2 (SD 2.1, Max 15.1)	15.7 (SD 11.6, Max 55.9)	332

Cognitive Effort and Efficiency in Translation Revision

the data (following Nakagawa & Schielzeth, 2013, but see also Gries, 2015). Collinearity was assessed by inspecting variance inflation factors for the predictors; all values were relatively low (<2), indicating that collinearity between predictors was not a problem. An overview of all predictor and dependent variables is given in Table 3.

The descriptive statistics in Table 1 show that the mean absolute task durations vary considerably from language to language and the variation between participants was also considerable, which is why we decided to use relative durations for the final revision phase – also because it is more informative and transferable than absolute durations: translators in real life typically translate texts which are much longer than the ~150 words used in this study. What may seem a small amount of time can become a considerable chunk of a translator’s workload. If scaled up to an 8-hour working day the final revision phase might actually take up to a maximum of between 2.8 and 4.4 hours, based on our data. We therefore used the dependent variable *% Revision* because we were interested in the relative amount of time translators spent on the final revision phase and considered the absolute time translators spent on the final revision phase less informative. The reason was also that after testing the same predictors used in Model 1 on the absolute time translators spent on the final revision, only *Total Fixation Duration (Revision)* and *Total Deletions (Revision)*, i.e., the total number of characters that have been deleted - this includes spaces and punctuation marks, but given that Translog does not allow formatting, it does not include keystrokes which contribute to formatting (see Table 3 for definitions of the variables) - remained significant. In other words, none of the predictors describing the behaviour during the drafting phase had a significant effect on the absolute time translators spent on the final revision phase, which is why we decided to use the relative values as dependent variable. We also used the *Total Task Duration* as a dependent variable because we wanted to find effects of online revision behaviour on the overall task duration - a longer overall task duration being indicative of a less efficient process. We chose predictors that indicated online revision behaviour and which could have an impact on translation quality and thus on final revision (see Table 3).

Table 3. Overview of variables used in final models including descriptions

Variable	Description	Indicator for	Unit
<i>% Concurrent ST Fixations (Drafting)</i>	Percentage of ST fixation during the drafting phase that took place concurrent to typing in relation to all ST fixations during drafting	Parallel ST reading while typing	%
<i>% Revision³</i>	Relative final revision phase duration compared to overall task duration	Phase distribution	%
<i>Total Task Duration</i>	Duration of the entire process including all three phases (task completion)	Translation speed	min
<i>Total Deletions (Drafting/Revision)</i>	Total deletions count during the drafting or final revision phase	Revision	n
<i>Total Fixation Duration (Revision)</i>	Total fixation duration during final revision phase	Time spent reading	min
<i>Total Scatter (Drafting)</i>	Total count of typing position changes within production units during drafting; Amount of continuous but non-linear text production (see below for a more detailed description)	Online revision	n
<i>Total Typing Pause Duration (Drafting)</i>	Total duration of pauses between continuous typing during drafting phase	Translation planning, problem solving, and/or revision	min

Model 1: Percentage Revision

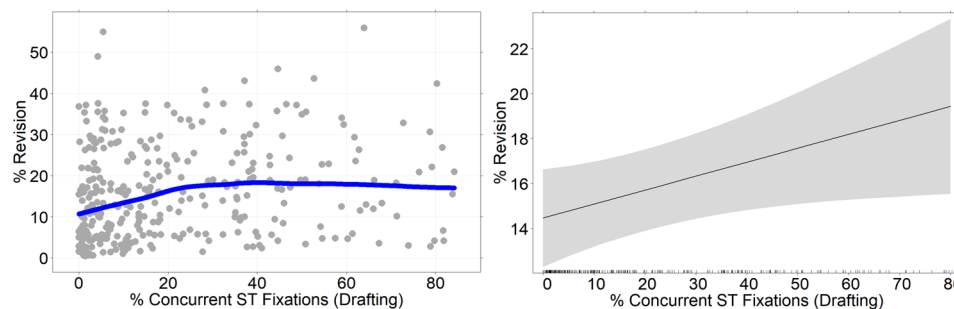
The dependent variable of the first model was the duration of the final revision phase relative to the total task duration (% Revision). The random variables (with random intercepts only) for both models were the same: participant, text and target language, given that we sampled 164 unique participants from six different languages not from the same language family and each participant translated three of the six English STs. As predictors, we included a number of variables from the drafting phase which indicate online revision. We were hoping to be able to predict which aspects of the behaviour during the drafting phase make the final revision phase longer or shorter – in relation to the total task duration.

While Figure 1 left shows the non-parametric curve for the effect of % *Concurrent ST Fixations* during the drafting phase on % *Revision*, Figure 1 right is a visualisation of the effect as described by Model 1.

The first predictor % *Concurrent ST Fixations (Drafting)* is explained as follows: Carl and Kay (2011) describe a measure called production unit. They find that pauses between two keystrokes of at least one second result in cognitively and linguistically plausible production units which also present cognitively and linguistically plausible translation units in regards to the typing behaviour. However, while translators type, they might also read either the ST or the TT. If translators read the TT while they type, it is highly likely that they are monitoring production and making sure that no typos or other mistakes occur. If translators read the ST while typing the TT, it is likely that they read either the words they translate or other relevant ST material – relevant to what they are currently typing or what they are about to type, since they might read ahead (Dragsted & Carl, 2013). In other words, concurrent TT reading and typing can be interpreted as – possibly monolingual – monitoring while concurrent ST reading and typing certainly involves the co-activation of two linguistic systems (source and target languages) and it also involves two more or less simultaneous processes: reading and writing. This kind of concurrent activity is therefore reminiscent of simultaneous interpretation in terms of co-activation of linguistic systems and simultaneity of reception and production processes.

To sum up, the predictor % *Concurrent ST Fixations (Drafting)* represents concurrent ST reading during production units which occurred during the drafting phase. It is expressed as a percentage of the total number of ST fixations and therefore also tells us something about the time spent reading the ST during a production unit. This predictor seems relevant for revision as it indicates split-attention between the ST and the TT. It can be described as a process in which the translator ensures that the TT corresponds to the ST in terms of, e.g., accuracy. The model shows that the effect of this variable is

Figure 1. Non-parametric curve and effect plot of the effect of the % *Concurrent ST fixations* during the drafting phase on % *Revision*



Cognitive Effort and Efficiency in Translation Revision

positive and significant at the 0.05 level: the more time translators spend reading the ST, the longer the final revision phase becomes (see Figure 1).

This effect can be interpreted in the following way: split attention with increased attention to the ST or, conversely, reduced attention to the TT during production decreases awareness of what has actually been produced in terms of TT units and increases the need to spend more time on the revision once a first draft of the ST is completed. In other words, the translator may intend to make sure that ST and TT correspond in terms of, e.g., accuracy, but split attention may make it more difficult to monitor TT production. It has to be stressed that the reading activity represented by this measure is concurrent to coherent typing, which is why attention is split between two concurrent tasks and two co-activated linguistic systems.

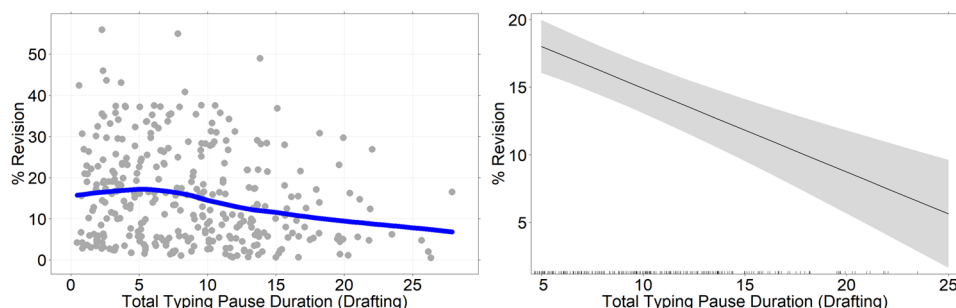
The second predictor *Total Typing Pause Duration (Drafting)* may also represent reading activity, but does not contain any typing activity, i.e., attention is not, or not to the same degree, split between tasks and linguistic systems.

The predictor *Total Typing Pause Duration (Drafting)*, here indicated in minutes, is the total pause duration in the drafting phase. Pause time is defined as the time before and after production units, i.e. continuous typing with pauses below one second. During this time, there are different possibilities of what participants might do – either they read the ST to plan their translation or the TT to check their translation and possibly plan revision, or they read both ST and TT to either plan translation or revision. They may also simply think about a problem and blankly look at the screen or somewhere else.

We also tested the effect of the total time spent reading the ST during drafting and the total time spent reading the TT during drafting in Model 1 – both had a strong, negative and significant effect on % Revision (total time spent reading the ST during drafting $t=-2.37$, $p < 0.05$ and total time spent reading the TT during drafting $t=-4.30$, $p < 0.001$). However, collinearity became problematic (variance inflation factors for these two predictors were > 2), which was not the case if *Total Typing Pause Duration (Drafting)* was included. In any case, no typing occurs during these pauses and while two linguistic systems might be activated to some degree, participants are not typing and reading concurrently. The effect of *Total Typing Pause Duration (Drafting)* was negative and highly significant (see Figure 2). In conjunction with the positive effect of concurrent ST reading, this effect suggests that the more time is spent on non-concurrent reading the less time is required or spent on revision of the final draft.

Keystrokes can occur sequentially in relation to the sequence of words in the TT or they may occur non-sequentially, i.e., a translator may type letters belonging to words in the sequence in which they

Figure 2. Non-parametric curve and effect plot of the effect of Total Typing Pause Duration during the drafting phase (in minutes) on % Revision



appear in the TT or a translator may insert letters in words which are separated by two or more words. This kind of non-sequential typing is captured by the measure *Scatter* described in Carl, Kay and Jensen (2010) and Carl et al. (2016). The larger this value the more often participants edit words which are separated by at least two other words. This measure within the drafting phase is highly likely to represent online revision as opposed to linear drafting given that already existing text is edited, presumably because there was a mistake or because the translator decides to change the existing draft for some other reason. The predictor *Total Scatter (Drafting)* represents the total number of times that editing during drafting production units occurred in a non-linear fashion, i.e., the more often a word which is separated by at least two other words has been edited the larger this value. The effect of this predictor was positive, such that the more often non-linear editing occurs during a production unit during the drafting phase, the longer the subsequent final revision phase. The effect was significant at the 0.05 level (see Figure 3).

The fact that the effect of *Total Scatter (Drafting)* was positive was surprising, since *Total Scatter* was thought to represent online revision and should therefore have a negative effect on % Revision. However, non-sequential typing in general might have a disruptive effect on a coherent text level representation in the translator's mind: the more often the translator jumps from the currently edited word to a more distant word, the more difficult it will be to represent the TT as a linear text and this might be the reason why subsequently the final revision phase needs to be longer so that it counterbalances the disruptive effect of editing discontinuous words.

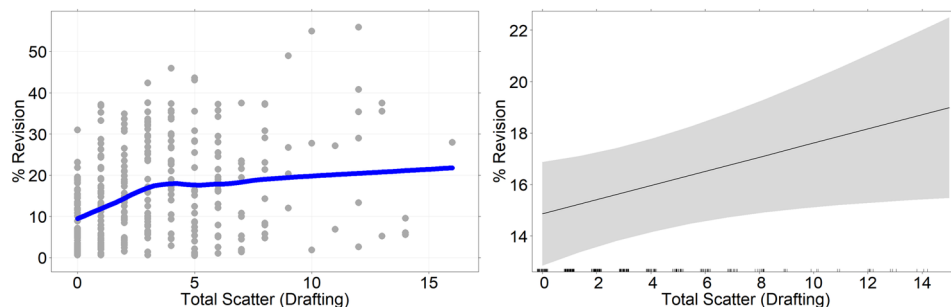
Similar to *Total Scatter (Drafting)*, the number of deletions that occur during a production unit are also indicative of online revisions, i.e., deletions which occur during a production unit may be caused by typos or other kinds of revisions. The effect of *Total Deletions (Drafting)*, visualized in Figure 4, was negative and significant at the 0.05 level.

However, unlike the effect of *Total Scatter (Drafting)*, the more revision (as represented by deletions) is carried out during the drafting phase, the less revision is necessary during the final revision phase (see Figure 4).

As control variables, we included the number of deletions which occurred during the final revision phase (*Total Deletions (Revision)*) and the total fixation duration during the final revision phase (*Total Fixation Duration (Revision)*) – irrespective of whether fixations were on the ST or the TT. Both had an expected strong, positive and highly significant effect (see Table 4).

Consequently, Model 1 provided a good fit (marginal $R^2=0.64$, conditional $R^2=0.84$). The AIC value for the null model was 2525 and including all the predictors, the AIC was reduced to 2168 for Model 1.

Figure 3. Non-parametric curve and effect plot of the effect of Total Scatter during production units in the drafting phase on % Revision



Cognitive Effort and Efficiency in Translation Revision

Figure 4. Non-parametric curve and effect plot of the effect of total number of deletions during production units in the drafting phase on % Revision

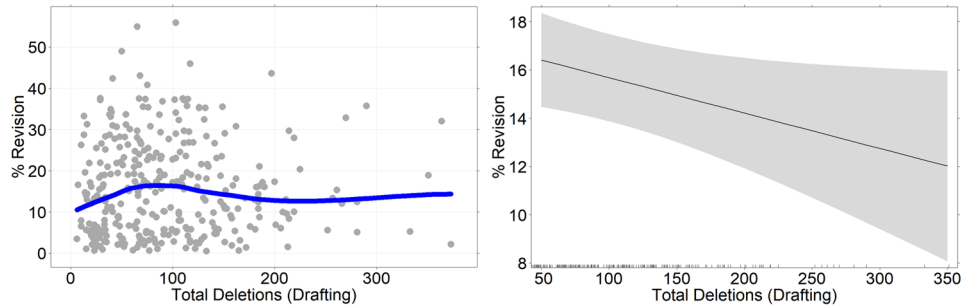


Table 4. Model 1. The effect of predictor and control variables on % Revision

Model 1	% Revision ~ Predictors + (1 Text) + (1 PartUnique) + (1 TL)			
	Sessions: 332; PartUnique: 158; TL: 6; Text: 6			
Predictors	β	Standard Error	<i>t</i>	<i>p</i>
% Concurrent ST Fixations (Drafting)	0.06	0.03	2.11	< 0.05 *...
Total Scatter (Drafting)	0.28	0.13	2.09	< 0.05 *...
Total Deletions (Drafting)	-0.01	0.01	-2.05	< 0.05 *...
Total Typing Pause Duration (Drafting)	-0.62	0.10	-5.67	< 0.001 ***
Total Deletions (Revision)	0.11	0.01	8.58	< 0.001 ***
Total Fixation Duration (Revision)	3.93	0.32	12.37	< 0.001 ***

In sum, Model 1 explains key variables which can predict how much time participants spend revising a first draft. Split attention and the non-linearity of the typing process have a positive effect – the more time participants spend reading in one language while typing in a different language and the more often they edit distant words, the more time they spend revising the initial draft. The more disrupted the text level representation of the first draft of the TT, the longer the final revision phase. The more online revision occurs in the form of deletions and the more time translators spend engaged in only reading and not typing, the shorter the final revision phase. The disruptive effect of concurrent reading and writing on text level representations becomes obvious when the strong negative effect of the total pause duration during the drafting phase is considered.

Model 1 thus explains why translators revise the first complete draft of the TT: text level representations are disrupted by concurrent processes and discontinuous editing. However, it is yet unclear whether these variables have an effect on efficiency as measured by total task time. Model 2 will attempt to test this.

Model 2: Fast and Slow Translators

The dependent variable for Model 2 was the total task completion time, i.e. *Task Duration*. When log transformed, the dependent variable *Task Duration* was normally distributed and so we ran the model with log transformed values. We decided to test the same predictors as in Model 1: % Concurrent ST Fixations (Drafting), Total Scatter (Drafting), Total Deletions (Drafting), Total Typing Pause Duration

(Drafting), Total Deletions (Revision), and Total ST Fixation Duration (Revision). Also, we included the same random variables *ParticipantUnique*, *Text*, and *TL*.

Interestingly, % Concurrent ST Fixations (Drafting) and Total Deletions (Drafting) had strong and highly significant effects in the opposite directions as compared to Model 1. While concurrent ST reading had a positive effect on the relative amount of end revision, Figure 5 shows that it seems to be more efficient timewise to read the ST concurrent to TT typing as it shortens the overall *Task Duration*. The effect of the % Concurrent ST Fixations (Drafting) on *Task Duration* was negative. The median value for task duration was 12.2 minutes.

The predictor *Total Deletions (Drafting)*, on the other hand, had a positive effect on *Task Duration*. The more deletions occur in the drafting phase, the longer it took the participants for the overall translation task. This is particularly surprising, given that the same predictor had a negative effect on % Revision. However, the tipping point for the total number of deletions during the drafting phase is relatively late: at around 115 deletions during the drafting phase, the total task time is likely to be longer than 12.2 minutes, i.e. above the median (see Figure 6).

The other two drafting predictors of Model 2 had similar effects as in Model 1. Both *Total Scatter (Drafting)* (Figure 7) as well as *Total Typing Pause Duration (Drafting)* (Figure 8) had a positive effect on the *Task Duration* which was not surprising. The effect of *Total Scatter (Drafting)* was significant at the 0.05 level and *Total Typing Pause Duration (Drafting)* was significant at the 0.001 level.

Figure 5. Non-parametric curve and effect plot of the effect of % Concurrent ST fixations during the drafting phase *Task Duration* (in minutes)

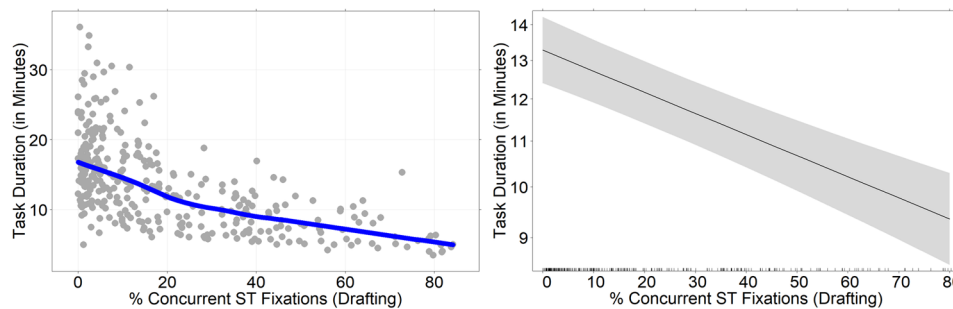
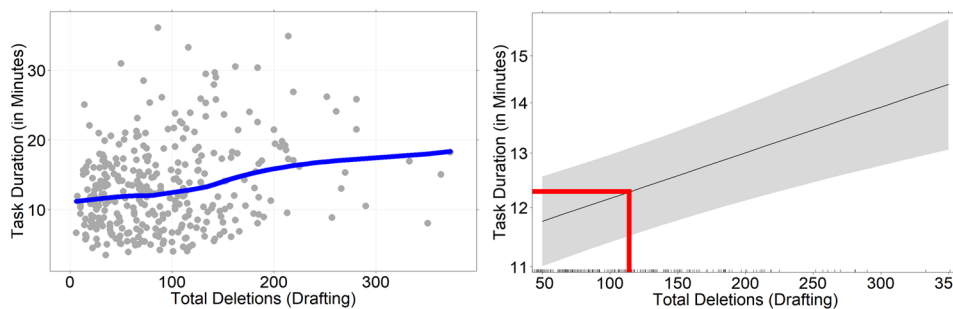


Figure 6. Non-parametric curve and effect plot of the effect of Total Drafting Phase Deletions on *Task Duration* (in minutes)



Cognitive Effort and Efficiency in Translation Revision

Figure 7. Non-parametric curve and effect plot of the effect of Total Scatter on the Task Duration (in minutes)

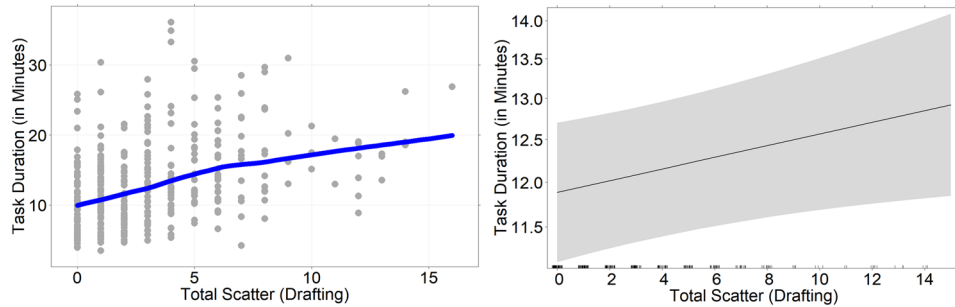
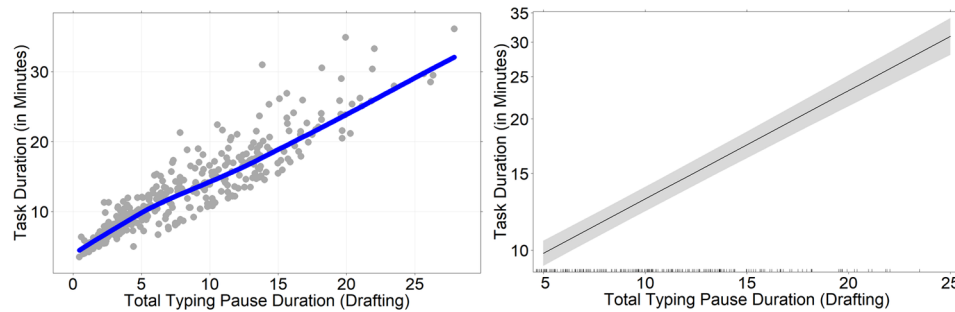


Figure 8. Non-parametric curve and effect plot of the effect of Typing Pause Duration in the drafting phase on the overall Task Duration (in minutes)



As Task Duration includes all pauses it is not surprising that the longer participants are not writing the longer they take in general irrespective of what they do during typing.

We also included the same two control variables from Model 2: *Total Deletions (Revision)* and *Total ST Fixations (Revision)*. Both had an expectedly strong, positive and highly significant effect (see Table 5). Model 2 provided a very good fit (marginal $R^2=0.89$, conditional $R^2=0.96$). The AIC for the null model was 2525 and was reduced to 2168 for Model 2 with all the predictors.

Table 5. Model 2. The effect of predictor and control variables from Model 1 on Task Duration

Model 2 Sessions: 332; PartUnique: 158; TL: 6; Text: 6		log(Task Duration) ~ Predictors + (1 Text) + (1 PartUnique) + (1 TL)			
Predictors	β	Standard Error	t	p	
% Concurrent ST Fixations (Drafting)	-0.004	0.001	-7.30	< 0.001 ***	
Total Scatter (Drafting)	0.005	0.003	2.17	< 0.05 **	
Total Deletions (Drafting)	0.001	0.0001	4.59	< 0.001 ***	
Total Typing Pause Duration (Drafting)	0.06	0.002	25.73	< 0.001 ***	
Total Deletions (Revision)	0.001	0.0002	5.70	< 0.001 ***	
Total ST Fixation Duration (Revision)	0.05	0.006	8.03	< 0.001 ***	

In sum, Model 2 adds to the findings from Model 1 and explains the same key variables which can predict how much time participants spend on the final revision phase. While split attention has a positive effect on the final revision phase it has a negative effect on overall task duration. The more time participants spend reading in one language while typing in a different language the more efficient is their drafting phase which makes up for the time they spend on final revision. However, both indicators for online revision – deletions and scatter – have a positive effect on *Task Duration*, in that participants take longer overall. A recommendation for translators that results from these two Models is thus, that the final revision is not only important in terms of quality control, but that online revision seems to be less efficient in terms of the overall task duration than an efficient drafting phase and a relatively longer final revision phase. On the basis of these findings, we wanted to find out whether the tested variables from Model 1 and 2 can be linked to participants’ professional status and experience as well as to their final translation quality. This will be discussed in the following section.

Interaction With Expertise and Quality

We tested the two presented models for interactions with expertise and we also annotated the German data with quality judgments. Table 6 gives an overview and descriptions of the expertise and quality variables that we included and tested in the model. While the metadata on participants’ experience and expertise was taken from the TPR-DB, the error annotations were performed by two independent annotators based on the quality assessment criteria suggested by Mertin (2006) which are explained in more detail in this section.

Expertise in our study was defined in four ways: *ExperienceYears*, as in the years of professional translator experience, *Status*, as a binomial variable based on years of experience (expert >2 years of professional experience), *Student*, as in student status (yes/no), and *TrainingYears*, as in the number of years that a participant had been trained as a translator. . We tested the first three of these variables in the two presented models. Unfortunately, none showed significant effects, so that interactions could not be tested either. Part of the problem seems to be the unbalanced dataset with far more students than professional translators as can be seen in Table 7. In addition, the metadata from the TPR-DB did not provide e.g. years of experience for all participants (see NAs). Also, there were students that had professional experience of >5 years and non-students with less than two years of experience. Metadata on *TrainingYears* was also collected but there was no further differentiation on what kind of training.

Table 6. Variables and their descriptions that were used to test effects and interactions with translation expertise and quality

Variable	Description	Indicator for	Scope	Unit
<i>Errors</i>	Total error count in final TT	Translation quality	Product	n
<i>ErrorsW</i>	Total error count including weighted errors in final TT, taking into consideration gravity of error	Translation quality (weighted)	Product	n
<i>ExperienceYears</i>	Years of professional experience as a translator	Translator status > expertise	Participant	years
<i>Status</i>	Experience	Translator status > expertise	Participant	yes/no
<i>Student</i>	Status whether participants were translation students or not	Student status > expertise	Participant	yes/no
<i>TrainingYears</i>	Years of professional translator training	Translator status > expertise	Participant	years

Cognitive Effort and Efficiency in Translation Revision

Table 7. Mean years of experience and training and count of sessions and participants per student status

Student Status	Mean Experience Years (Years)	Mean Training Years (Years)	Novice Sessions	Expert Sessions	All Sessions	Participants
Student: yes	1 (SD 3, Max 20, NA:59)	5 (SD 1.9, Max 6, NA:104)	146	35	240	117
Student: no (Professionals)	10 (SD 7.9, Max 33, NA:0)	3 (SD 2.2, Max 7, NA:32)	2	91	93	47
Total	2 (SD 7.2, Max 33, NA:59)	4 (SD 2, Max 7, NA:136)	207	126	333	164

For future expertise judgement, we suggest an annotation for the participants that takes into consideration not only student status but also years of formal translator training and years of professional translator experience. This data needs to be collected for all participants across studies. We assume that this discrepancy influenced the test and while no conclusions can be drawn on expertise interaction this still proves to be an interesting variable for further studies. The fact that we did not find any significant effects with expertise is surprising, given that it could be expected that professionals are generally more efficient due to their continued practice and superior skill.

Quality in this study was measured based on the quality assessment criteria suggested by Mertin (2006). We used all but one criteria from the two criteria categories *formal-linguistic* and *translation-relevant* criteria (no ST errors needed to be corrected in the translation). Errors included the following of which four were weighted (indicated by a 2): *typing error* (Tp), *punctuation* (I2), *spelling* (R2), *grammar* (Gr2), *mistranslation* (S2), *terminology* (W), *omission* (A), *redundant information* (RI), *stylistic error* (St), *consistency error* (K), *missing text coherence* (Tk) and *localization error* (Lk). All 46 TTs from the German subset were annotated independently by two anonymous annotators and finally by one of the authors. Unfortunately, we were only able to annotate the German translations which are only 13.2% of all sessions recorded. To extend the underlying research, an annotation for the sessions of the other five studies is necessary which will hopefully provide a better base for the statistical analysis. For the quality assessment the errors were counted and added to the dataset. We tested the model with two translation/revision quality variables: the absolute error count (*Errors*) and weighted error count (*ErrorsW*). We did not test for separate error types as at this point it did not seem feasible given the small German dataset.

Table 8 shows that per text there are only 8 sessions (for text 1 only 7) and that there is not much variance between the 6 different texts with regard to the number of errors. Text 2 seems to be slightly out of line and participants showed the highest error count and weighted error count. Since we have one data point per participant and thus not an extremely large dataset in total, it is not surprising that the much smaller German dataset does not show any effect with regard to translation quality measured in errors in the final TT. With regard to the efficiency of online revision vs. final revision, interim translations, too, i.e. drafts at the end of the drafting phase, would have to be annotated according to the error matrix and compared to the final TT. Expanding this study with an annotation of all six translated languages and extending annotation to interim translations could prove a promising path to find out more about what types of revision prove to be not only efficient with regard to translation speed but also to translation quality.

Table 8. Descriptive summary statistics for quality annotation of the German TTs with mean and count for absolute and weighted errors per text

Text	Mean Errors (Count)	Mean ErrorsW (Count)	Errors (Sum)	ErrorsW (Sum)	Number Sessions
1	3 (SD 2, Max 6)	6 (SD 3.4, Max 10)	23	39	7
2	6 (SD 1.9, Max 9)	9 (SD 2.7, Max 13)	50	68	8
3	4 (SD 2.9, Max 10)	6 (SD 4.8, Max 17)	35	51	8
4	3 (SD 2.8, Max 8)	5 (SD 4.3, Max 12)	26	38	8
5	3 (SD 1.6, Max 5)	5 (SD 2.4, Max 8)	25	40	8
6	5 (SD 2.0, Max 6)	8 (SD 2.6, Max 10)	39	61	8
Total	4 (SD 2.4, Max 10)	6 (SD 3.6, Max 17)	198	297	47

CONCLUSION

In summary, we can say that while split attention, as captured by concurrent ST reading during a production unit in the drafting phase, has a positive effect on the relative time translators spend reading the first complete draft during the final revision phase, this effect is paled by the strong effect this very same measure has on total task duration: already once translators spend around 15% of the total number of fixations during a production unit on ST reading, they belong to the group of overall faster and thus more efficient translators. A 2-3% decrease in the revision time at around 135 deletions during the drafting phase is gained in terms of the relative amount of time spent revising the complete first draft (see Figure 2). This is not a major gain in terms of total time. However, the same number of deletions or more during the drafting phase puts translators into the slow group, if total task time is considered.

It is thus possible to say that the most efficient process involves a large degree of concurrent reading and writing and few deletions during the drafting phase. The efficiency gains in terms of relative revision duration achieved by avoiding discontinuous typing, by making a larger number of deletions, pausing for longer amounts of time and engaging in less concurrent reading and writing are outweighed by the gains in total task time by doing the exact opposite.

One obvious limitation of this study is that we have only been able to annotate the final TTs in terms of quality for the German data and any claims regarding the effect of the behaviour during translation on the quality of the final TTs must, therefore, remain rather limited for the moment, although this aspect would be very informative and useful and will be considered in future studies. These findings could result in the formulation of concrete guidelines on efficient revision strategies. Their application might have an effect on the quality of professional translation processes on the one hand and on revision training in translation courses on the other. Future research could involve a long-term study on how the implementation of such guidelines will improve the efficiency of revision processes and their impact on translation quality assurance in general.

REFERENCES

- Alves, F., & Vale, D. C. (2011). On drafting and revision in translation: A corpus linguistics oriented analysis of translation process data. *Translation: Corpora, Computation, Cognition. Special Issue on Parallel Corpora: Annotation, Exploitation, Evaluation, 1*(1), 105–122.
- Bartoń, K. (2009). *MuMIn: Multi-model inference*. R Package Version 0.12.2/r18. Retrieved from <https://CRAN.R-project.org/package=MuMIn>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using {lme4}. *Journal of Statistical Software, 67*(1), 1–48. doi:10.18637/jss.v067.i01
- Carl, M., & Kay, M. (2011). Gazing and typing activities during translation: A comparative study of translation units of professional and student translators. *Meta: Translators' Journal, 56*(4), 952–975.
- Carl, M., Kay, M., & Jensen, K. T. H. (2010). Long distance revisions in drafting and post-editing. *Proceedings of CICLing-2010*. Retrieved from http://static-curis.ku.dk/portal/files/131448111/2010_Long_Distance_Revisions_CarlKayJensen.pdf
- Carl, M., Schaeffer, M. J., & Bangalore, S. (2016). The CRITT translation process research database. In M. Carl, S. Bangalore, & M. Schaeffer (Eds.), *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB. New Frontiers in Translation Studies* (pp. 13–54). Cham: Springer International Publishing.
- Dragsted, B., & Carl, M. (2013). Towards a classification of translation styles based on eye-racking and keylogging data. *Journal of Writing Research, 5*(1), 133–158. doi:10.1007/978-3-319-20358-4_2
- Drugan, J. (2013). *Quality in professional translation: Assessment and improvement*. London: Bloomsbury Academic.
- Englund Dimitrova, B. (2005). *Expertise and explicitation in the translation process*. Amsterdam: John Benjamins. doi:10.1075/btl.64
- Fox, J. (2003). Effect Displays in {R} for Generalised Linear Models. *Journal of Statistical Software, 8*(15), 1–27. doi:10.18637/jss.v008.i15
- Gouadec, D. (2007). *Translation as a profession*. Amsterdam: John Benjamins. doi:10.1075/btl.73
- Gries, S. T. (2015). Some current quantitative problems in corpus linguistics and a sketch of some solutions. *Language and Linguistics (Taipei), 16*(1), 93–117. doi:10.1177/1606822X14556606
- Ipsen, A. H., & Dam, H. V. (2016). Translation revision: Correlating revision procedure and error detection. *Hermes: Journal of Language and Communication in Business, 55*, 143–156.
- ISO 17100 (2015). *Translation services: Requirements for translation services (ISO 17100:2015)*; English. Berlin: Beuth.
- ISO 9001 (2015). *Quality management systems: Requirements. Chapter 8 (ISO 9001:2015)*; German and English versions. Berlin: Beuth.

Jakobsen, A. L. (2002). Translation drafting by professional translators and by translation students. In G. Hansen (Ed.), *Empirical Translation Studies: process and product* (pp. 191–204). Copenhagen: Samfundslitteratur.

Jakobsen, A. L. (2003). Effects of think aloud on translation speed, revision and segmentation. In F. Alves (Ed.), *Triangulating Translation: Perspectives in Process Oriented Research*, 2 (pp. 69–95). Amsterdam: John Benjamins; doi:10.1075/btl.45.08jak

Künzli, A. (2005). What principles guide translation revision? A combined product and process study. In I. Kemble (Ed.), *Translation norms: What is ‘Norma’ in the translation profession? Proceedings of the Conference Held on 13th November 2004 in Portsmouth* (pp. 31–43). Portsmouth, UK: University of Portsmouth, School of Languages and Area Studies.

Künzli, A. (2006). Translation revision: A study of the performance of ten professional translators revising a technical text. In M. Gotti & S. Šarčević (Eds.), *Insights into Specialized Translation* (pp. 195–214). Bern: Peter Lang.

Künzli, A. (2009). Qualität in der Übersetzungsrevision: Eine empirische Studie. In H. Kalverkämper & L. Schippel (Eds.), *Translation zwischen Text und Welt: Translationswissenschaft als historische Disziplin zwischen Moderne und Zukunft* (pp. 291–303). Berlin: Frank & Timme.

Künzli, A. (2014). Die Übersetzungsrevision: Begriffsklärungen, Forschungsstand, Forschungsdesiderate. *trans-kom*, 7(1), 1–29. Retrieved from http://www.trans-kom.eu/bd07nr01/trans-kom_07_01_01_Kuenzli_Revision.20140606.pdf

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). {lmerTest} Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1–26. doi:10.18637/jss.v082.i13

Malkiel, B. (2009). From Antonia to my Antonia: Tracking self-corrections with Translog. In S. Göpferich, A. L. Jakobsen, & I. M. Mees (Eds.), *Behind the mind: Methods, models and results in translation process research*. Frederiksberg: Samfundslitteratur.

Mertin, E. (2006). *Prozessorientiertes Qualitätsmanagement im Dienstleistungsbereich Übersetzen. Leipziger Studien Zur Angewandten Linguistik Und Translatologie*, 2. Frankfurt am Main: Peter Lang.

Mossop, B. (2007). Empirical studies of revision: What we know and need to know. *JosTrans: The Journal of Specialised Translation*, 8, 5–20.

Mossop, B. (2014). *Revising and editing for translators* (3rd ed.). New York: Routledge.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142. doi:10.1111/j.2041-210x.2012.00261.x

R Development Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: The R Foundation for Statistical Computing.

Rasmussen, K. W., & Schjoldager, A. (2011). Revising translations: A survey of revision policies in Danish translation companies. *JosTrans: The Journal of Specialised Translation*, 15, 87–120.

Cognitive Effort and Efficiency in Translation Revision

Robert, I., Remael, A., & Ureel, J. J. J. (2016). Towards a model of translation revision competence. *The Interpreter and Translator Trainer*, 11(1), 1–19. doi:10.1080/1750399X.2016.1198183

Robert, I. S. (2008). *Translation revision procedures: An explorative study*. Retrieved from <http://www.arts.kuleuven.be/cetra/papers/files/robert.pdf>

Robert, I. S. (2012). *La révision en traduction: les procédures de révision et leur impact sur le produit et le processus de révision* (Unpublished doctoral dissertation). University of Antwerp, Antwerpen.

Robert, I. S. (2014). Investigating the problem-solving strategies of revisers through triangulation: An exploratory study. *Translation and Interpreting Studies*, 9(1), 88–108. doi:10.1075/tis.9.1.05rob

Shih, C. (2006). Revision from translators' point of view: An interview study. *Target: International Journal on Translation Studies*, 18(2), 295–312. doi:10.1075/target.18.2.05shi

Shih, C. (2013). Translators' "end-revision" processing patterns and maxims: A think-aloud protocol study. *Arab World English Journal*, 4(3), 31–44.

Shreve, G. M., Angelone, E., & Lacruz, I. (2014). Efficacy of screen recording in the other-revision of translations: episodic memory and event models. *MonTi: Monografías de traducción e interpretación*, 1, 225–245.

ENDNOTES

¹ Rasmussen and Schjoldager (2011, p. 87) also list the very small number of empirical studies that had been carried out up to the year 2011.

² sites.google.com/site/centretranslationinnovation/tpr-db

³ *%Revision* refers to the time spent on revision after the first draft has been produced relative to the total time spent on producing the final version per session, while *Mean %Revision* in Table 2 refers to the mean relative time spent on revision across all sessions.

Section 4

Learning Translation Quality Assessment

Chapter 11

Constructing Standards in Communities: Tutors' and Students' Perceptions of Assessment Practices on an MA Translation Course

Elsa Huertas-Barros
University of Westminster, UK

Juliet Vine
University of Westminster, UK

ABSTRACT

Assessment practices on translation programs provide a valuable lens through which to view current understandings about the nature of translation pedagogy. In the context of competence-based training, the last decade has seen the proliferation of assessment instruments aimed at enhancing students' learning by prioritising competence development and the translation process. Using the University of Westminster as a case study, the authors have sought to provide a clearer insight into the current understandings of translation and assessment practices on the MA Translation courses in the light of the current debates in translation pedagogy. The authors undertook a two-pronged approach by surveying not only the tutors, but also the students. This chapter will present and analyse the findings of the two surveys on assessment practices using the framework of the six tenets of good assessment practice set out by the Higher Education Academy (HEA) and, in particular, assess to what extent assessment literacy has been developed.

INTRODUCTION

Translation training in universities has proliferated in the last 20 years. In 1998, Caminade and Pym (quoted in Kelly, 2005, p. 8) listed 250 university programmes in 60 countries as an example of the explosion of courses on offer. In a recent research project, the authors found there were 27 universities in

DOI: 10.4018/978-1-5225-5225-3.ch011

the UK alone offering translation training for professional contexts (Huertas Barros & Vine, 2018). The field of translation studies, and along with it translation pedagogy research, has flourished. Much of this translation pedagogy research has been focused on trying to analyse the complex set of skills, attitudes and knowledge which combine to enable translation. This analysis has led to a variety of models (e.g. Kelly, 2002, 2005, 2007; PACTE, 2003, 2005; González Davies, 2004; EMT Expert Group, 2009; Kiraly, 2013) and these in turn have been conceptualised as translator competence. The components of translator competence have then been used to establish the learning outcomes (LOs) for core translation modules.

At the same time as researchers and academics have been honing their models and debating the component aspects of translator competence, universities have recruited tutors to teach translation. Many of these tutors fit the Holroyd's (2000) profile for teachers¹ throughout the HE sector, of being experts in their field of knowledge, but having 'craft' knowledge of assessment, i.e. knowledge gained 'from experience of being assessed and assessing others, any training needed is likely to be minimal' (2000, p. 34). Rust *et al.* use the term 'tacit' knowledge, a term coined by Polanyi in 1958 and later adopted by Nonaka (1991), describing it as 'Deeply rooted in action and often in an individual's commitment to a profession, tacit knowledge consists partly of technical skills based on professional experience, and in a more cognitive dimension, in our ingrained mental models, beliefs and perspectives' (Rust, Price, & O'Donovan, 2003, p. 152). They go on to say that tacit knowledge 'underpin[s] the established normative, "connoisseur" model of assessment—illustrated by the phrase "I cannot describe it, but I know a good piece of work when I see it"' (Rust *et al.*, 2003, p. 152). However, Paran believes that 'intuitions and beliefs are not reliable when complex issues such as teaching and learning are concerned. This is where a research-oriented or evidence-based approach to teaching comes into play' (Paran 2017, p. 501).

In its paper *The EMT Translator Trainer Profile Competences of the trainer in translation* (2013), the European Master's in Translation (EMT) recognised the importance of considering the skills and aptitudes that translator trainers should 'possess or develop' (EMT Expert Group, 2013, p. 1). The paper includes the disparate sets of professional skills that trainers bring to the training and mentions that the trainers are not specialists in education and suggests that 'depending on national regulations, teacher training, either as a formal qualification or additional, bespoke, or specialised teacher training relevant to the course(s) taught is highly desirable' (EMT Expert Group, 2013, p. 1). However, this remains an aspiration not a requirement.

Given the ubiquity of the concept of translator competence in the literature on translation pedagogy and translation studies, including, for example, the inclusion of a model of translator competence in the EMT Expert Group (2009), and given the lack of explicit training in issues relating to translation pedagogy of many tutors in university courses, the authors' main objective is to investigate the current understandings of translation and assessment practices on an MA Translation course. In the light of the data gathered, the authors aim to analyse the extent to which assessment practices being used at present relate to the current debates on translation pedagogy and also the wider debates on assessment in Higher Education (HE). By surveying not only the tutors but also the students on MA Translation courses at the University of Westminster, the authors were able to not only triangulate the tutors' responses, but also to recognise in their research an important aspect of learning and teaching in relation to assessment. In particular, it is only through explicit engagement of both tutors and students in discussions that clear understandings about the nature and purpose of assessment can be created.

ASSESSMENT IN HIGHER EDUCATION

Assessment practices on translation programmes provide a valuable lens through which to view current understandings about the nature of translation pedagogy. In the introduction to her book *A Handbook for Translator Trainers*, Kelly states that the debate on ‘training translators has only rarely transcended our self-imposed disciplinary borders to take account of the wider issues of higher education and training’ (2005, p. 2). The situation does not seem to have changed much over the last decade despite the recent attention given to assessment in HE. Indeed, in undertaking research for this paper, the authors also found few references that relate the assessment practices in translator education to the wider HE debates on assessment. The authors believe that viewing the discipline-specific practices in this wider context could provide valuable fresh perspectives on the debates within translation studies on how to assess translator competence. The authors also agree with Kelly:

Not only does Translation Studies have things to learn from these general trends, but — given its interest in training — it also has in turn quite a lot to offer current debates on changes in higher education. (Kelly, 2007, p. 129)

It has been recognised that across the HE sector, there is a problem with assessment (Boud, 2010; Assessment Standards Knowledge exchange [ASKe], n.d; HEA, 2012; Price, Rust, O’Donovan, Hindley, & Bryant, 2012). In the 2016 National Student Survey (HEFCE, 2016), assessment remained the area of greatest student dissatisfaction. While students had a satisfaction rating for the teaching on their courses of 87%, the rating for assessment was 74%. This is the lowest rating for all the areas surveyed and well below the benchmark for acceptability set by the universities at 85%. Assessment has been given the lowest rating in all the surveys since 2010. Each of the papers quoted above sets out a list of premises or proposals which if followed would improve assessment practice. Boud and Associates (2010) make seven proposals, the first of which states students should be ‘inducted into the assessment practices and cultures of higher education’ and that ‘[s]tudents and teachers become responsible partners’. The UK ASKe (n.d.) in its position paper states that ‘assessment is central as it frames student learning’ and as its second premise proposes ‘assessment is more effective when students understand the assessment process’ (p. 2).

In 2012, the UK Higher Education Academy (HEA) published ‘*A Marked Improvement: Transforming Assessment in Higher Education*’, which following on from the work of ASKe, sets out six tenets for assessment (HEA, 2012, p.21):

1. Promoting assessment for learning. This includes considering the balance of formative and summative assessment, where formative assessment provides opportunities for practising.
2. Developing assessment fit for purpose. Assessment should align with the stated LOs; assessment should not prioritise reliability over validity, for example exams should not be chosen as the assessment primarily for reasons of reliability but because they do reflect the LOs.
3. Recognising that assessment lacks precision. This means recognising that not all learning is accessible to assessment and that trying to measure all the LOs can constrain the assessment process. When assessing complex LOs, professional judgements can be used. However, these professional judgements need to be constructed and agreed on within appropriate communities.

4. Constructing standards in communities. There should be dialogue between tutors and students to ensure there is a shared understanding of assessment and a mutual trust is established.
5. Integrating assessment literacy into course design. This means that within the teaching process, the subject of assessment is explicitly addressed. Both the students and tutors involved should understand the skills required in the assessment process. Understanding these skills will allow students to be autonomous learners.
6. Ensuring professional judgements are reliable, i.e. academic judgement is a 'holistic judgement rather than mechanistic processes. Academic, disciplinary and professional communities should set up opportunities and processes, such as meetings, workshops and groups to regularly share examples and discuss assessment standards.'

Although it is more than five years since the HEA published its paper encouraging universities to transform their assessment practices, there continue to be significant problems with assessment. The HEA published a report in 2016 of a summit into the changes in assessment practice and admitted that 'assessment remains the area perhaps least effectively engaged by efforts at change' (Elkington, 2016). The report suggests that the key to improvement in assessment is assessment literacy and the importance of providing staff with opportunities to develop their understanding of assessment literacy.

As the brief outline of the six tenets shows, there is considerable overlap between them. The tenet of 'assessment literacy' could be construed of as an umbrella tenet under which all the other tenets must be gathered in order for it to be fully realised. Smith, Worsfold, Davies, Fisher, and McPhail (2013) conceptualise assessment literacy as the capacity of students to develop a 'multi-dimensional' set of skills and understandings with respect to assessment. These were divided into three areas. Firstly, an understanding of 'the purpose of assessment and how it connects with their learning trajectory' (Smith et al., 2013, p. 45). Secondly, students need 'to be aware of the processes of assessment' (Smith et al., 2013, p. 45), including an understanding of the criteria and standards, and how these processes affect their ability to respond to the tasks set. Thirdly, the students need to be given opportunities to practise evaluating their own work so that they 'can learn to identify what is good about their work and what could be improved' (Smith et al., 2013, p. 45). Smith et al. focus on the assessment literacy of students, but as Elkington (2016, p. 4), states raising awareness of assessment literacy should 'integrate[s] developments' within the staff and student bodies.

In order to develop assessment literacy, both staff and students need to learn about and reflect on all the tenets and how these tenets will impact on their engagement in the assessment process. Assessment literacy ensures transparency of assessment processes and thereby confidence on behalf of the tutors and trust on behalf of the students that the process is both valid and reliable. For institutions to be transparent, those responsible for assessment will need to reflect on their tacit understandings of the process and in shared communities of practice expose them to scrutiny and together construct shared agreed understandings. As Elkington (2016) states the 'key first step will be encouraging all relevant staff to review their own assessment practice' (p. 15) which will involve turning their craft or tacit knowledge into shared standards within their communities of practice.

ASSESSING TRANSLATOR COMPETENCE

In the context of competence-based training, the last decade has seen the proliferation of assessment instruments aimed at enhancing students' learning by prioritising competence development and the translation process. In the light of the current debates in translation pedagogy and using the University of Westminster as a case study, the authors have sought to provide a clearer insight into the current understandings of translation and assessment practices on the MA Translation courses. The integration of competence-based training in translator education has led to a gradual shift in focus from the product of translation to the process of translating. This translation process has been broken down into several interrelated areas of competence or sub-competences that translation scholars have conceptualised in several models of translator competence (e.g. Kelly, 2002, 2005, 2007; PACTE, 2003, 2005; EMT Expert Group, 2009). However, despite the general consensus on the main skills or sub-competences, many of the existing models 'do not describe its operationalisation in sufficient detail (its components and indicators, [...] performance levels, etc.) to allow for the planning of teaching' (Galán-Mañas & Hurtado Albir, 2015, p. 68)².

The shift in focus has influenced the way translator educators perceive the function of assessment, with the last two decades witnessing innovations in assessment and feedback in translation across Europe. 'Traditional' assessment practices to judge students' performance (i.e. summative product assessment only) are being questioned and there is an increasing number of research papers promoting new forms of assessments. These include diagnostic and formative assessment, summative assessment with a formative function, peer and self-assessment, translation commentaries, reflective diaries, students' portfolios, or translation process recordings of students or professionals while they are translating (Hurtado Albir, 1999/2003, 2007, 2015a, b; González Davies, 2004; Kelly, 2005; Way, 2008, 2014; Galán Mañas & Hurtado Albir, 2015; Huertas Barros & Vine, 2015, 2016, 2018; Lisaité et al., 2016). However, given the complexity of assessment in which subjectivity plays an important part, there is a need for further research (González Davies, 2004, p. 31) in the search for more objective, valid and reliable product- and process-oriented assessment instruments.

The fundamental function of assessment in the teaching and learning process is witnessed by the number of dedicated edited volumes and papers addressing translator competence assessment and assessment issues in translation (Schäffner & Adab, 2000; Martínez Melis & Hurtado Albir, 2001; Tsagari & van Deemter, 2013). Drawing on the dichotomy of bridging the gap between pedagogical and professional assessment, some research also explores process-oriented and product-oriented evaluation approaches (Orlando, 2011, 2012). In the context of competence-based training, recent research suggests a range of competence assessment procedures, instruments and tasks for diagnostic, formative and summative assessment in translator training (Galán Mañas & Hurtado Albir, 2015).

Contributions to summative assessment include some proposals of frameworks for this type of assessment (Beeby, 2000, p. 185-198; Adab, 2000, p. 215-228; McAlester, 2000, p. 229-243), as well as empirical case studies at BA (Pavani, 2016) and MA levels (Klimkowski & Klimkowska, 2012; Huertas Barros & Vine, 2016, 2018). In the last decade, translation scholars have also devoted special attention to formative assessment of translation processes, with some proposals on self-assessment, self-criticism and decision-making frameworks (Way, 2008, 2009, 2014; Prieto Ramos, 2014). Massey and Ehrensberger-Dow (2012, 2013) analyse alternatives to traditional product assessment and suggest assessment methods

that focus on assessing translation processes in laboratory-based and workplace projects. This research in translator competence assessment shows increased acceptance of competence-based understanding of translation as well as increased awareness of pedagogical issues among translation scholars, including assessment.

EMPIRICAL CASE STUDY: TUTORS' AND STUDENTS' PERCEPTIONS OF ASSESSMENT PRACTICES ON AN MA TRANSLATION COURSE

This section will therefore present and analyse the findings of the two surveys informing the comparative case study of assessment practices on the core translation modules at the University of Westminster. The case study aims to provide a clearer insight into the current understandings of translation practice and assessment practices on the MA Translation courses. More specifically, the authors aim to investigate the following three main areas: 1) If and to what extent the concept of competence-based understanding of translation has been adopted by tutors and students; 2) How and where the two groups' perceptions on assessment practices converge and diverge; 3) How and in what ways this understanding of translation affects beliefs and attitudes to assessment practices. This section will present and analyse the findings of the two surveys using the framework of the six tenets set out by the HEA and, in particular, assess to what extent assessment literacy has been developed.

Research Methodology

The authors' case study draws on research methodologies successfully used in empirical studies in translator education (e.g. Saldanha & O' Brien, 2013; Huertas Barros, 2013; Huertas Barros & Vine, 2016, 2018; Huertas Barros & Buendía Castro, 2018). The case study presented in this paper is the first phase of a wider research project into assessment practices on MA Translation courses offered in the UK (Huertas Barros & Vine, 2018). This case study consisted of three stages: 1) a preliminary survey conducted with MA translation tutors (n=16) teaching on the core translation modules (Huertas Barros & Vine, 2016), 2) a revamped assessment grid in response to tutors' feedback, and 3) a survey carried out with MA translation students (n=53) where the changes to the assessment practices and refined feedback sheet were tested. Both courses, i.e. the MA in Specialised Translation and the MA in Translation and Interpreting, were included in the case study.

The survey of tutors (n=16) was carried out in 2015 and the results were used as a basis for consultation and discussion amongst the teaching staff on the university's MA courses. Translation tutors actively engaged in dialogue about assessment standards and processes. As a consequence of this consultation, changes to the assessment practices and feedback sheet were made in response to the tutors' feedback and implemented the subsequent academic year (i.e. 2016-2017) (see section Refined Assessment Criteria). The MA students were then surveyed at the end of that academic year (n=53). Therefore, the student survey not only allowed for triangulation with the original findings, but was also able to discover if the changes sought by the tutors were supported by the students. The sites of divergence in the perspectives of the two groups can indicate areas where a shared understanding of and mutual trust in assessment may need to be strengthened.

The questionnaires were designed with predominately closed questions, but some open-ended questions were also included to provide qualitative information about the rationale behind participants' responses.

Constructing Standards in Communities

Both questionnaires were designed using GoogleForms and followed a similar structure to allow for triangulation of the information. Both surveys included the following sections: Section A) General details: qualifications and experience; Section B) Tutors/Students' perceptions of assessment; and Section C) MA Translation assessment criteria and feedback sheet used on the core translation modules.

Analysis and Discussion of the Findings of the Comparative Case Study

The structure of the survey lies at the basis of this section, which is divided into four subsections. The first subsection gives some background information about the MA in Specialised Translation / MA in Translation and Interpreting courses and the participants. In the second subsection, the authors provide an insight into tutors' and students' perceptions of assessment practices. Some of the aspects explored include perceptions of and attitudes to: 1) The extent to which it is important to assess the different areas of translator competence in pieces of summative assessment; 2) The implications of assessing the translation product or the translation process; 3) The extent to which exam performance is considered an accurate representation of the students' typical translation competence; 4) The areas of competence which are most accurately measured by an exam; 5) The assessment instruments and tasks used on the core translation modules; 6) Proposed changes to the weighting of such assessment tasks. The third subsection analyses the refined assessment criteria and feedback sheet used on the core translation modules. The final subsection discusses tutors' and students' perceptions of the levels of objectivity versus subjectivity involved in marking.

Background Information About the MA Translation Courses and the Participants

The University of Westminster is a particularly interesting academic institution in which to carry out a case study, because the present MA courses were developed out of a long history of professional training in translation. It was established as the first polytechnic in the UK and offered the first specialised training course in Translation (Pedrola, 1999, p.19). This focus on training for the professions, which polytechnics promoted, has again become the focus of debate about reforms in HE. The university has a long tradition of using professional translators as tutors on its core translation modules and these modules were designed to offer practical training reflected in the emphasis on professional experience in those teaching on the courses. However, in the last decade and certainly since the introduction of the EMT, there has been a growing awareness of models of translator competence and changes have occurred at course level such as the introduction of new modules, including CAT, Audiovisual Translation and Project Management. This reflects the need to offer training in a wide range of the competences identified.

The responses to the questions on the qualifications and experience of the tutors surveyed showed that the majority (87.5%) had an academic qualification in Translation (i.e. Diploma in Translation, BA, MA and/or PhD) and 68.8% were members of a professional association. Fifty per cent of the tutors had been teaching on translation modules for 10 years or more, and 18.8% had been part of the teaching team for between 6-9 years. Only 12.5% of the staff who completed the survey were relatively new to the course (i.e. 1-2 years). Virtually all tutors (93.8%) had worked in the translation industry (either as freelance translators, in-house translator and/or as project managers) for between 6-9 years (26.7%) or for 10 or more years (73.3%). In most cases (73.3%), tutors combined their teaching with freelancing work and over two thirds had attended workshops and/or training courses on teaching translation skills, pedagogical approaches or assessment practices. As for students who completed the survey, 26.4% had previously

completed a BA in Translation and 62.3% had experience working as translators (i.e. volunteer translator - 60.6%; freelance translator - 39.4%; in-house translator - 27.3%; other, including internships - 9.1%).

The core translation modules cover a wide range of specialised technical and institutional texts. Before the first survey was conducted, the core translation modules were assessed by a combination of two pieces of coursework and one end-of-year examination. Translation coursework consisted of two written translation assignments (350-500 words) worth 30% of the overall module mark (i.e. 15% each). The examination consisted of a 3-hour, open book, translation on a computer (650-750 words). Formative assessments (i.e. translation assignments and other translation-related tasks) were also provided to students on a regular basis.

Tutors' and Students' Perceptions of Summative Assessment

To gain insight into tutors' and students' awareness of translator competence, the first question in this section asked both tutors and students to rank the different areas of translator competence in terms of how important it is to assess them in the pieces of summative assessment used on the core translation modules. For this purpose, the authors provided a definition of each area of translator competence in both surveys based on the definitions suggested by Kelly in her model of translator competence (2005: 32-33). As shown in Table 1, the data demonstrates a fairly strong correlation (correlation coefficient of 0.7385) between tutors' and students' answers. Students' understanding of translation and the concept of translator competences predominantly comes from the tutors, which may explain the correlation between their answers. However, 62.3% of the students had had some translation experience as translators beyond the MA course. A follow-up multiple choice question revealed that 60.6% of these students had worked as volunteer translators, 39.4% had worked as freelance translators and 27.3% had had professional in-house translation experience, so their responses can also be seen as their own evaluation.

One of the points which varies is the importance the students attach to intercultural competence compared to the tutors (2nd and 5th respectively). This could be a result of the students' own experience with many students coming to England from other countries to study and thus being aware of the different cultures in a more immediate manner. It could also be related to the students' perceptions of translation and their previous translation experience. Furthermore, one of the option modules offered on the course

Table 1. The extent to which it is important to assess the following competences in pieces of summative assessment⁴

Translator Competences	Tutors' Ranking	Students' Ranking
Language competence	1 st	1 st
Intercultural competence	5 th	2 nd
Professional competence	2 nd	3 rd
Information mining	3 rd	4 th
Thematic competence	4 th	4 th
Strategic competence	3 rd	5 th
Technological competence	6 th	6 th
Social / interpersonal competence	7 th	7 th

Constructing Standards in Communities

focuses on intercultural competence exclusively and the fact that it is offered may influence students' perceptions about the significance of this competence.

The second main difference was related to the ranking of the strategic competence, which tutors ranked higher than students (3rd vs. 5th respectively). The analysis of the tutors' responses revealed this sub-competence had the widest variation in rankings, which reflected the fact that tutors did not all believe this component was as critical as the translator competence models suggest (e.g. Kelly, 2002, 2005; PACTE, 2003, 2005; EMT Expert Group, 2009). Despite the survey having provided a definition of this competence³, it is possible the respondents were still unclear about what exactly it encompasses. This lack of clarity may have been reflected in the tutors' responses, or perhaps the tutors felt that strategic competence is not directly assessed, but is reflected in the products. Since tutors are transmitting their understanding of translator competence to students, it is possible that they also transmit their lack of clarity, which may explain why students also placed strategic competence towards the bottom of the ranking. The lack of consistency in the rating of the strategic competence may also reflect the issue raised by Galán-Mañas and Hurtado Albir (2015, p. 68) that translator competence has not been 'operationalised' and as such it is difficult for the tutors and by extension the students to have clear understandings of some areas of competence.

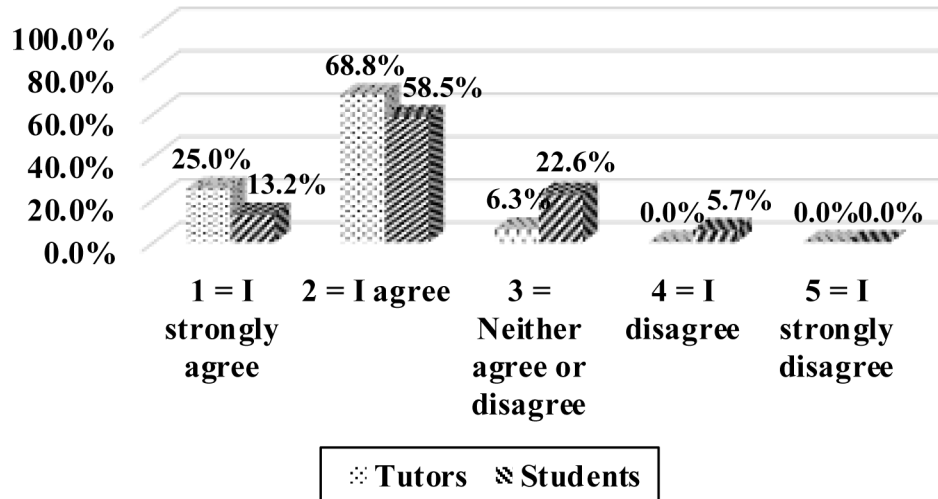
The ranking given to the translator competences mirrors those competences which are explicitly addressed in the LOs of the modules. The exception to this was Language Competence, a competence that it is assumed students have on admission to the courses and so not identified as an LO. As de Groot (2000) states 'not all components that can be distinguished in a criterion task will need to be trained because a number of them may be mastered already at the onset of training' (p. 54). These responses were also broadly in agreement with what tutors thought was measured most accurately by an examination (see Figure 5), which evidences the fact that tutors felt that those competences which are most important to assess are indeed the ones which are assessed. On the other hand, some tutors raised the need for complementary assessment methods that consider other important aspects of translator competence, especially social/interpersonal competence.

In response to the question asking if assessing a translation (i.e. the product of the translation process) was an accurate reflection of the students' competence, there is a tendency for the tutors to be more positive in their evaluation of the assessments used on the core translation modules (see Figure 1). Except for one tutor, they all (+93%) agree or strongly agree that assessing the product of the translation process is an accurate representation of students' translator competence compared to 71.7% of students.

A significantly higher number of students (28.3%) chose an ambivalent response or disagreed, while only 6.3% of the tutors expressed ambivalence regarding the validity of the product as representative of students' competence. The lower levels of positivity about the assessment methods could reflect the fact that students do not feel clear about the assessment practice and why it is valid. The HEA tenet 'constructing standards in communities' (HEA, 2012) suggests that if students are more involved in the discussions of the assessment practice, they will have higher levels of trust in the assessment and feel more positive about its use, so this discrepancy could indicate that there is a need for increased discussion and a more collaborative approach to constructing assessments.

In most cases, tutors and students provided a very similar range of qualitative responses to justify their views. Most tutors believe that the product 'is the only sensible way to assess students', although some tutors felt that the product does not give an account of certain competences such as the social/interpersonal competence or technological competence. Students felt that the product is what future clients and employers will assess, and it is 'the overall summary of my entire work.' However, the tutors' responses

Figure 1. Do you agree that assessing translations (i.e. the PRODUCT of the translation process) accurately reflects students' translator competences?



were more emphatically positive about the ability of the product to assess translator competence. The students raised the issue of the subjective nature of the marking, which according to them would not necessarily reflect the translator competence: ‘Not always. Some assessment has been very subjective and focused more on the way a tutor liked to write’⁵. One student was also concerned that the tutors may assess without having fully researched the subject area and, consequently, being less informed. These concerns with the competence of the assessors were not raised by the tutors themselves. The fact that both the students and tutors used the same language to describe the issues with assessing the product, shows that there is some shared understanding of assessment and, from this, it can be inferred that there is a degree of ‘assessment literacy’ (HEA, 2012). However, the concerns about the validity again show that more needs to be done to ensure mutual trust i.e. students and tutors need to engage in a dialogue about standards and assessment procedures (‘Constructing standards in communities’, HEA, 2012).

Both groups also agreed that the translation product is the most easily managed method of assessing translator competence within an academic institution, with 75.1% of tutors and 77.3% of students agreeing or strongly agreeing on this point (see Figure 2). The tutors’ responses were more emphatic, since nearly twice as many respondents (i.e. 43.8% of tutors compared to 24.5% of students) strongly agreed. A higher proportion of students (18.9% compared to 12.5% of tutors) chose an ambivalent response.

When asked if assessing the translation process is an important aspect of assessing translator competence within an academic institution (see Figure 3), both tutors and students provided the same range of responses (nearly 70% of tutors and students agreed or strongly agreed). The qualitative data gathered from both groups reflect the importance of the translation process during the training and the value of assessing the process as part of formative assessment, with some students stating that ‘it is the only way to understand the weaknesses and strengths in the translation.’ Some tutors concurred with students in this regard, and stated that ‘without consistent good process there cannot be consistent good products’ and that ‘it [the translation process] offers explicit learning opportunities and allows students to reflect on strategy.’ Those students who disagreed felt that ‘as long as the product is good, the process you use to get there is not particularly relevant or important.’ In a quarter of the students’ responses, they

Constructing Standards in Communities

Figure 2. Do you agree that assessing a translation *PRODUCT* is the most easily managed method of assessing translator competence within an academic institution?

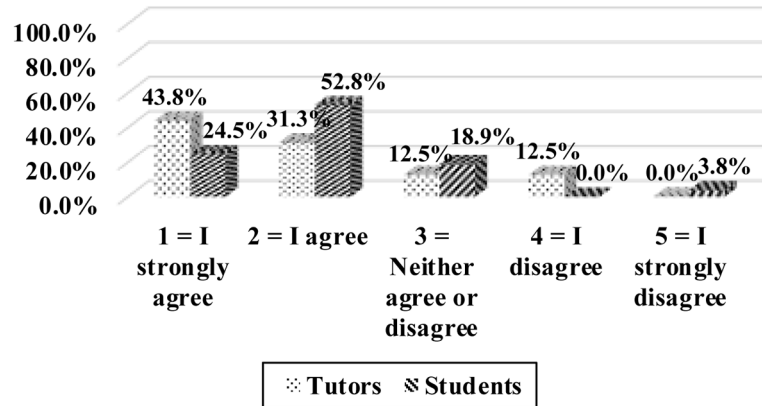
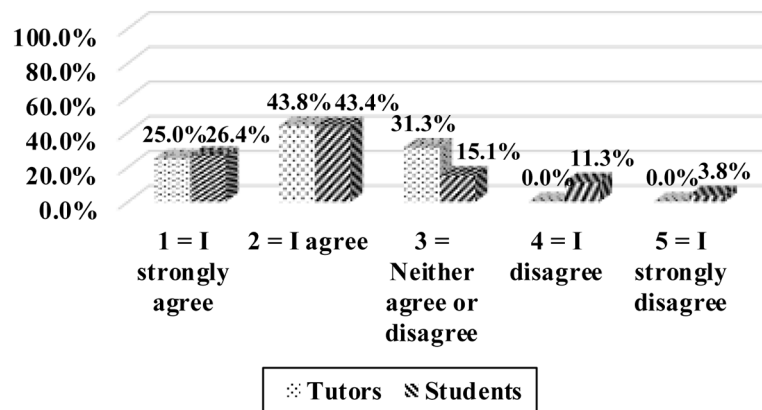


Figure 3. Do you agree that assessing the translation *PROCESS* is an important aspect of assessing translator competence within an academic institution?



stated that the process should be taught but it should not be assessed, in other words, they believed the process should be discussed in class as part of the ongoing formative feedback/ assessment they received throughout the module, but only the final product should be assessed since ‘this is how our work will be judged as professionals.’

While both groups’ responses highlight the importance of ‘formative assessment’ as an indication of how students are performing at each stage, only 12% of tutors and 8% of students explicitly stated the process should be an important aspect of the summative assessment. The students suggested that taking the process into account may allow tutors to see the reasoning behind choices which they may not agree with. The qualitative responses showed a clear understanding of the distinction between product and process which in turn indicates that some aspects of assessment literacy are already integrated into the core translation modules.

As shown in Figure 4, there is general agreement between both tutors and students that exam performance is not necessarily an accurate representation of students’ typical translator competence. Both

Figure 4. Do think your exam performance is an accurate representation of your (students') typical translator competence?

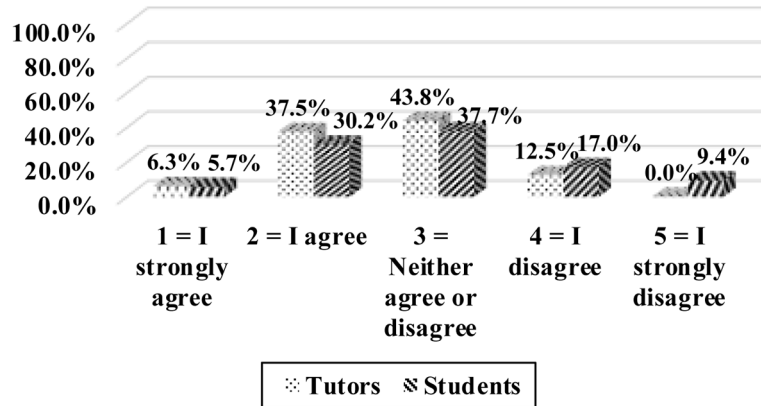
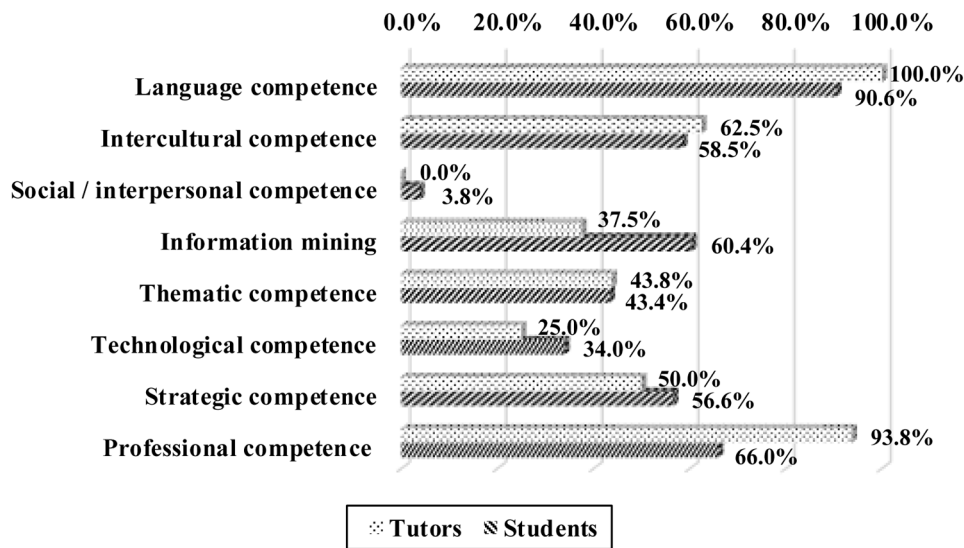


Figure 5. Which of the following competences are most accurately measured by an examination?



groups are less certain about the accuracy of exams to represent translator competence than they were of the ‘product’ in general. Only 5.7% of the students fully agreed with the statement that ‘exam performance is an accurate representation of [their] translator competence’ and 30.2% agreed, giving a total of 35.9% agreeing to some degree. At the same time, more than one in three students (37.7%) neither agreed nor disagreed, and 26.4% disagreed or strongly disagreed with this statement. The students’ qualitative comments were even more cautious in their assessment of the accuracy of exams to reflect their translator competence than was evidenced in the quantitative information. Of the students who had not strongly agreed with the proposition in the qualitative question, 60% did not agree that the exam was a true reflection of their typical translator competence. These students felt that exams were not representative of the conditions under which they would normally translate, e.g. additional pressure and time constraints, and no opportunity to put the translation aside before proof reading: ‘As a freelance I would have the option

Constructing Standards in Communities

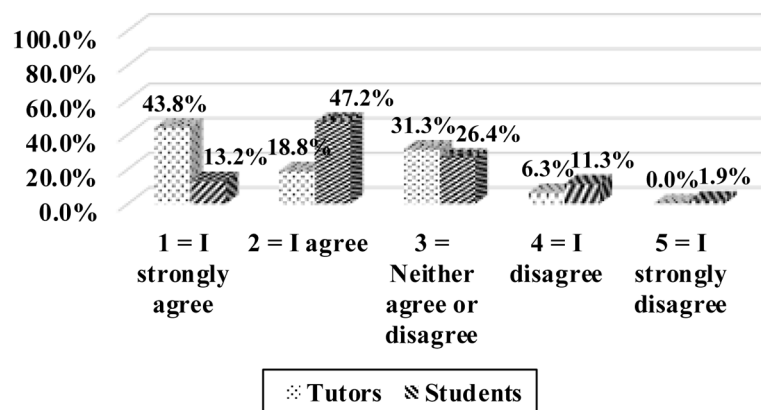
of turning down a text which one does not have in an exam, 3 hours solid sat at a computer does not fit in with my working pattern, nor does it provide ample time to do additional research where needed.’ The inability to put the translation aside and continue to work on it at a later stage with a fresh mind was mentioned by several students. Of the remaining students, 25% agreed exams did represent their typical translator competence, but they had reservations in line with the students who disagreed, and only 10% of those who wrote comments agreed without reservation, while 5% were unsure.

With regard to the tutors’ perceptions, 43.8% agreed that exam performance was an accurate representation of translator competence, while 56.3% of tutors expressed ambivalence or disagreement with this statement (see Figure 4). This ambivalence about exams is also reflected in the response to questions about the assessment weightings (see Figure 6). Some tutors pointed out that some students produce excellent coursework but do not perform equally well in the exam. One tutor added that ‘exam conditions are generally not representative of the translation work done by a professional, whereas coursework is.’

This ambivalence in regard to the use of exams from both the tutors’ and students’ perspective indicates that there is a need to re-assess the validity and reliability of this assessment method. A re-assessment of the use of exams is in line with the HEAs second tenet, ensuring assessment is fit for purpose, which states ‘there needs to be recognition of the difficulties inherent in marking systems, and the imbalance between validity and reliability needs to also be addressed through an increased emphasis on assessment validity (HEA, 2012, p. 19).

Despite the ambivalence reflected in the previous question, tutors and students believe that language and professional competence are the competences which are most accurately measured by an exam (language competence – tutors [100%], students [90.6%]; professional competence – tutors [93.8%], students [66%]) (see Figure 5). As shown in Table 1, both tutors and students rated these two competences as part of the top three competences to be assessed in the pieces of summative assessment. According to both groups, exams can also reflect students’ intercultural competence (tutors [62.5%], students [58.5%]). However, respondents agree that exams are not indicative of students’ social / interpersonal competence, and do not measure very accurately other areas such as technological or thematic competences. Nearly twice as many students (60.4% compared to 37.5% of tutors) believe that exams reflect their ability to

Figure 6. Do you agree that the assessment instruments and tasks used on the core translation modules (i.e. two pieces of coursework and one exam) can accurately measure the level of translator competence attained by students?



search for information (i.e. information mining). This mismatch in perceptions could be due to the fact that students may feel the research process carried out prior to the examination, when the exam topic is announced, makes use of their information mining competence and is therefore also assessed by the exam.

As can be seen in Figure 6, the quantitative data shows that tutors are more positive than the students about the accuracy of the assessment methods (i.e. two pieces of translation coursework and one exam) to accurately measure the level of translator competence attained by students, since nearly half (43.8%) strongly agreed, whereas 13.2% of students strongly agreed. However, a similar percentage of tutors (62.6%) and students (60.4%) generally agree with this statement, and there is also strong resemblance in the number of tutors and students giving ambivalent responses (tutors – 31.3%; students – 26.4%). Those tutors who did not strongly agree provided many of the same reasons as the students' for being concerned about the ability of coursework and exams involving translation only to fully reflect the whole range of translator competences. To solve this problem, these tutors suggested the inclusion of alternative assessment instruments that reflect the learning process (e.g. attaching to the translation a short commentary or annotations in which students justify their decision-making) and that are representative of translator competences which are generally overlooked (i.e. incorporating a group task to be able to assess social / interpersonal competence).

There were 35 qualitative responses from the students explaining why they thought the assessment was or was not an accurate reflection of their translator competence, 12 of whom did not make explicit in their answers if they agreed or disagreed with the question, thereby leaving 23 clear responses. Of these, eight students agreed that the assessment pattern did accurately reflect their translator competence, six students completely disagreed, two students were very positive about the coursework but very strongly against the use of exams, and the rest of the students were in agreement with the assessment pattern but had reservations such as the weighting given to the exam and whether the assessment assessed all competences.

Students referred to the same issues to argue two contradictory points of view. For example, some students argued that since the exam and coursework were two very different assessments, this meant that they could not be used together to give an accurate account of the students' ability. However, other students said that because the two assessment types were so different, it meant they were able to assess different and complementary competences and so gave a more accurate account. This pattern of resorting to the same issues to argue opposite points of view continued as some students felt the assessment instruments were not an accurate representation of their continuous development as translators, whereas others felt that they in fact allowed the assessment of different stages of the learning process. Moreover, some students did not see the exam as a valid assessment method as it did not reflect the situation in the professional world (i.e. a 3-hour translation) while other students felt that the examination pressure reflected the professional context and a realistic translation brief tests what students could achieve in the working world. As did several of the tutors, some students stated the translation coursework and exam do not assess all translator competences. The students' responses made confident and accurate use of the concepts underpinning translator competence and the dichotomies in their views reflect the debates in translator education literature, thus demonstrating that translator competences are well integrated into student understanding of translation. It also shows that students have an understanding of 'the purpose of assessment and how it connects with their learning trajectory', one of Smith et al.'s assessment literacy indicators (Smith et al., 2013, p. 45).

Despite most tutors agreeing that the assessment methods used accurately measure students' level of translator competence, the percentage of tutors who agreed with the weightings allocated to the pieces

Constructing Standards in Communities

of summative assessment on the core translation modules (i.e. translation coursework 1 – 15%, translation coursework 2 – 15%, exam 70%) was identical to the percentage of tutors who disagreed (i.e. 44%). Tutors felt the weighting allocated to the translation exam was excessively high and more importance should be attached to the coursework elements. They also felt coursework mirrors more accurately real-life translation assignments where translators have access to all resources and take time to go through various stages of the translation process, while exam conditions and the attendant time limitations do not always allow students to perform to the best of their abilities. The results obtained from the survey of tutors were presented and discussed in breakout-groups at a staff workshop on curriculum assessment and enhancement. The staff decided to reduce the weighting attached to the exam to 60% and to introduce a progressive weighting for the two pieces of coursework, giving greater weighting to the second piece of coursework to reflect students' learning curve (i.e. translation coursework 1 – 15%, translation coursework 2 – 25%, exam – 60%). In the survey of students, 83% of respondents strongly supported the changes (see Figure 7).

Refined Assessment Criteria and Feedback Used on the Core Translation Modules

The final section of the survey focused on the MA Translation assessment criteria and feedback sheet used on the core translation modules. The tutors' survey results (see Figure 8) showed that while 81.3% of tutors felt the assessment criteria used on the core translation modules were clear to them, 44% would add or remove certain elements to refine the translation feedback sheet.

The original feedback sheet included the following criteria: comprehension, accuracy, readability (style/register), terminology/lexis, grammar/syntax, spelling/punctuation, and presentation. Tutors felt that accuracy (66.7%) and readability (55.6%) were the categories requiring further clarifications or refinements. The survey and the subsequent attendant course team discussions held at the curriculum assessment and enhancement staff workshop (see previous section) resulted in valuable revisions to the translation feedback sheet used for formative and summative assignments on all core translation modules. A revised rubric providing a more detailed definition of each criterion as well as examples of different issues included under each criterion were introduced to ensure that more focused and transparent feedback was provided. The revisions included a clarification of the criteria 'comprehension' and 'accuracy' to specify the corresponding stage of the translation process (i.e. 'comprehension' [decoding of ST]; 'accuracy' [encoding of TT]). The criteria 'readability (style/register)' was refined to exclusively include

Figure 7. Students' view on the proposed changes to the weighting of assessment tasks on the core translation modules

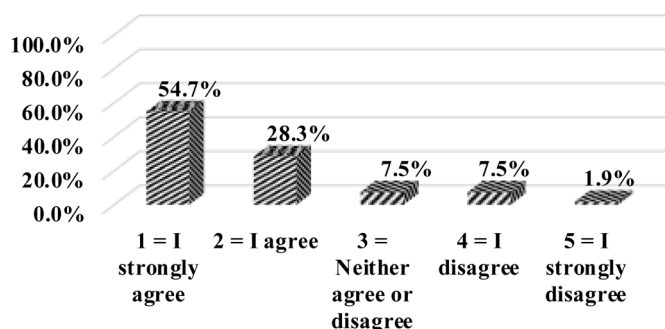
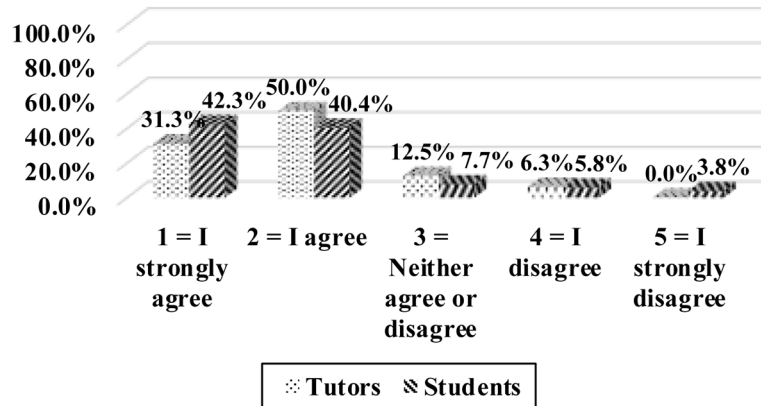


Figure 8. Tutors' and students' views on the clarity of the assessment criteria and parameters used on the core translation modules



stylistic aspects of a translation (i.e. register was removed from this category), and the criteria originally entitled ‘presentation’ was also removed from the feedback sheet. Instead, two new assessment categories (i.e. ‘compliance with brief’ and ‘preparation for final delivery’) were added to articulate more fully the professional context in which learning was situated. These two new categories accounted for register and presentation respectively as examples of the type of issues included under each criterion. These refinements were tested with the new cohort of students with positive results. A total of 82.7% students felt the assessment criteria used in the translation modules were clearly defined for them as students and 88.2% would not implement any further changes following the foregoing refinements. The initial tutors’ survey, including consultations on how the assessment criteria were and should be understood, then revising and sharing the refined criteria and feedback sheet was an important first step in raising assessment literacy with the staff group. This process also follows the one suggested by the HEA in its sixth tenet, ensuring professional judgements are reliable, by giving the staff group the opportunity to share and discuss academic standards.

The revamped criteria and feedback sheet were subsequently tested with students. The students’ opinions were gathered via the student survey and further discussion at course committee level, both of which provide a valuable first stage in engaging students in a process of constructing standards in communities.

Tutors’ and Students’ Perceptions of the Levels of Objectivity vs. Subjectivity Involved in Marking

Both surveys included questions about what the respondents felt was the balance between objectivity and subjectivity involved in assessment both in an ideal situation and in practice. When the authors designed the questions on objectivity and subjectivity, particularly in the context of the criteria employed in assessments and assessment in general, they had in mind a model in which objectivity was achievable when an aspect of the translation could be construed as ‘right’ or ‘wrong’. These would be the aspects which would lend themselves to being marked on a marking grid. The authors also recognised that some aspects of the translation could not be given a clear right/ wrong evaluation and these aspects were subjective as opposed to objective. However, the authors’ understanding of subjective assessment was assessment which

Constructing Standards in Communities

relied on professional judgement. This professional judgement is formed of ‘tacit’ knowledge (Nonaka, 1991) that has gone through ‘a sharing of experience–socialisation processes’ (Rust et al., 2003, p.152) involving discussion with other professionals in informal and formal mentoring, in marking moderation sessions or in staff development workshops, so that individual understanding becomes standardised.

In response to the question on the ideal balance between objectivity and subjectivity, over 81% of students either strongly agreed or agreed that marking should be completely objective (see Figure 9). However, when tutors were asked the same question, 43.5% felt marking should be completely objective or felt the balance should be 10%:90% subjective/objective (12.5% and 31.3% respectively). A further 25% felt the balance should be 20%:80% subjective/objective and nearly 20% of tutors felt that balance should be 50%:50% subjective/objective. These figures show there is a substantial mismatch between the students and tutors in their understandings of what are acceptable levels of subjectivity in assessment, which can be explained in the light of the different interpretation tutors and students seem to have of such concepts. The qualitative responses (see below) seem to indicate tutors refer to ‘professional experience/ judgement’, whereas students seem to take ‘subjective’ to be the tutor’s personal opinion or preference.

This disparity between tutors and students is found when both groups were asked about what they considered was the actual balance in practice (see Figure 10). Here both groups felt that the level of objectivity was lower in practice than it should be in an ideal situation. None of the tutors felt that their marking was completely or even 10%:90% subjective/objective. The students agreed that no marking was in practice completely objective, but over a quarter felt that the marking was 10%:90% subjective/objective. Half of the tutors believed that their translation marking was 20%:80% subjective/objective and another quarter believed that their marking was 30%:70% subjective/objective, which left the final quarter believing their assessment was 40% to 60% subjective.

The students had the expectation that assessment should be and is more objective than the tutors believe it should be or is. Given the conception of subjective explained in the opening paragraph, it is not surprising that the tutors, even in an ideal situation, were comfortable with levels of subjective evaluation in the assessment process ranging between 30-60%. However, as the students’ responses below show, they have a very different conceptualisation of ‘subjective,’ one which would lead them to see subjectivity as a weakness in the assessment. And this would undermine the students’ perceptions of

Figure 9. Students’ views on the requirement of a level of objectivity of a 100% and the actual balance in practice

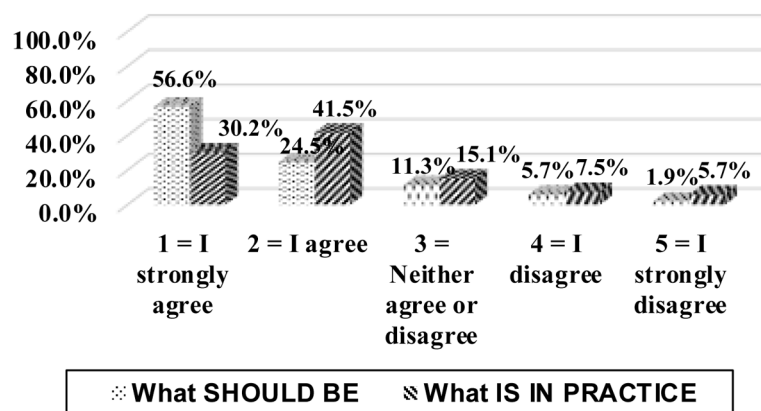
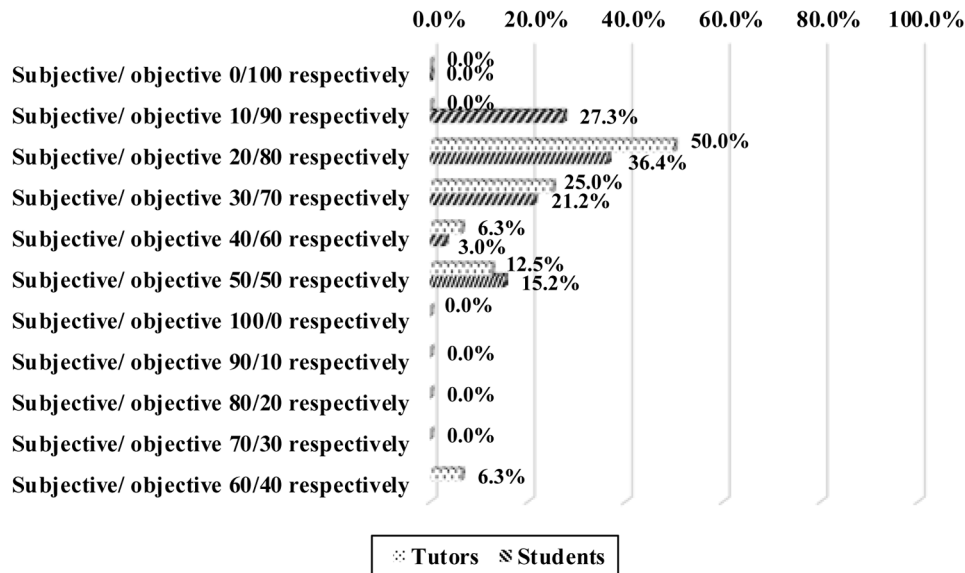


Figure 10. Tutors' and students' perception of the balance between 'subjective' and 'objective' in the assessment practices on the core translation modules



the reliability of the final marks and, therefore, the students felt that ideally there should be much lower levels of subjectivity.

The students provided 20 reasons for believing that marking was not completely objective. Many of the students recognised that it is the nature of the assessment that the marking is unlikely to be completely objective and some subjectivity is 'inevitable'. The reasons the students gave could be divided into four categories: tutor preference, differences (i.e. discrepancies) between tutors, tutor error and anonymity. Half of the reasons given relate to the issue of tutor preference. The students felt that 'markers have specific likes and dislikes that can influence them in their final assessment of the text' and that 'every teacher/person has a personal point of view about language', so that the 'marking process will inevitably have much to do with a teacher's taste in language and style'. There is a sense in the students' responses that these likes and dislikes are arbitrary in that they refer only to the individual and do not relate to or respond to wider professional standards. One student commented 'On more than one occasion, a teacher demonstrated that what they thought was good on one occasion, they had then changed their minds about on another. So, in that respect, the marking process cannot necessarily be all that objective, as much as one would hope the intentions of the process to be.' There does not seem to be a recognition that the same translation solution to the same problem might not be appropriate in different translation contexts. Students felt that the tutors' preferences were excessively rigid and did not allow for the recognition that alternative versions were also acceptable. One student wrote 'Some teachers are very good at marking objectively and accepting that there can be many different but equally good versions of the TT. However, many others mark you down if you do not completely comply with what they think is the appropriate style/tone and correct terminology.' One student felt that the criteria used were themselves inherently subjective, commenting 'I feel that while the assessment criteria are very useful in helping the marker to give an objective evaluation of the end product, it is almost impossible for them to be completely

Constructing Standards in Communities

objective.’ In particular, this student felt that criteria such as ‘readability’, ‘compliance with brief’ and ‘lexis’ may allow for personal preference in terms of register and choice of lexis.

The other categories of reasons included different tutors giving different marks. Three students commented that there was a discrepancy between tutors on the same module feeling that this resulted in very different marks. Two students raised the issue of students who are native speakers of the source language having a clearer understanding of the grammar and lexis than the tutors, but that the tutors were not ‘listening’ and merely ‘agreeing to differ’. In classes where native and non-native speakers of the target text language submit coursework together, one student indicated that the tutors ‘preferred’ the target text produced by native speakers. In cases where native and non-native speakers are in the same class and submitting the same coursework, there are two different marking criteria to account for the distinction in main language or second language translation, but from the student’s comment it is clear that there is some concern that this differentiation is not taking place in practice.

A quarter of the reasons given for questioning the objectivity of the assessment relate to anonymity of marking. This was surprising to the authors who had initially only considered the issue of subjectivity in terms of applying criteria to texts rather than in terms of who is being marked. All coursework and exams are marked anonymously, but issues raised by the students show that they do not believe that the system is entirely anonymous. One student noted ‘I feel that sometimes names are not fully removed from the documents or Turnitin⁶ does not guarantee complete privacy and it does not help if you are identified as an English native speaker or someone with English as a second language.’ One student felt ‘I think over the course of a term lecturers get to know the style of their students and can probably guess sometimes who wrote which text, it could only be sort of objective if an independent third person marks and assesses the assignments’. These answers are revealing as they demonstrate that there is need for the systems used to guarantee anonymity to be clearly explained and that students have an opportunity to express their concerns and have them addressed.

The divergence in responses is revealing in terms of the lack of trust the students have in various aspects of the assessment process. These findings show that there is a need for the tutors to continue to build their shared understanding of professional judgements to ensure there is parity and transparency in the standards they apply. The tutors also need to be able to share their understanding of the nature of professional judgement so that students are assured that it is more than mere individual preference. These discussions would contribute to an increase in levels of assessment literacy, as would more discussion on the criteria that the students are assessed by. This process has been started but as Elkington points out ‘Like other literacies assessment literacy is slowly learned and never complete’ (2016, p. 5).

CONCLUSION

The widespread introduction of translator competence models in translator education demonstrates that, as a discipline, translation has engaged in debates about the most effective forms of training, which includes debates on assessing. This is also testified by the wide range of research on assessing translator competence. The case study presented here reveals that this discipline-wide debate is also informing both tutors’ assessment practices and students’ evaluation of these practices. In their responses to questions on assessment, the tutors and students showed they generally had a clear understanding of translator competence models and their responses echoed the debates on the appropriateness of assessing certain

competences, product or process, the validity of a final exam and other assessment tasks, and how objective marking can be.

As a discipline, translation has always had a clear understanding of the links between university education and the professional translation industry. This understanding has meant that assessment practices have been linked to real-world practices thus ensuring that assessment reflects the demands of the professional contexts. Translator educators within universities and the translation profession have debated for some time the skills that need to be developed and assessed. This means that translation as a discipline has already implemented many aspects of the HEA tenets for good assessment practice. For example, translation training recognises that translator competence can only be achieved through practice, which means that formative assessment is integrated on to translation modules, thus ensuring that summative assessment is preceded by formative assessment. Given the clear LOs of modules using translator competence as their theoretical underpinning, the assessment and LOs can be clearly aligned. In terms of the first two assessment tenets i.e. ‘promoting assessment for learning’ and ‘developing assessment fit for purpose’ the authors’ concur with Kelly that ‘[w]e are, then, in an ideal position to become involved in more general movements towards the renovation and innovation of university education in general; (...) and it is surely our responsibility as trainers to share our experience’ (Kelly, 2007, p. 130).

However, the authors’ research has revealed concern about current assessment practices amongst both tutors and students. It has also revealed differences in levels of satisfaction with and trust of the assessment instruments. These findings show that there are areas where the HEA tenets could be more fully employed to create assessment practices which both tutors and students feel are valid and reliable. The most important area to address is ‘integrating assessment literacy into the course design’ and by doing this the other tenets such as ‘constructing standards in communities’, ‘recognising that assessment lacks precision’ and ‘ensuring the reliability of professional standards’ will also become embedded in assessment practice. The work carried out with the tutors, using the results of the tutor survey, discussing the assessment practices and criteria and making adjustments to the assessment weighting and the criteria is an important step in developing assessment literacy among the staff and constructing standards in communities. Consulting students’ perspective about these refinements can also be seen as a way of incorporating these two tenets.

Given the relatively small sample size we cannot generalise these results to larger populations. However, the results can indicate potential trends that may contribute to and/or inform the practice of assessment in translator education and future research on theoretical discussions and follow-up empirical studies. The case study may be considered an informed suggestion for best practice in translator education, i.e. assessment literacy can be integrated more effectively into course design to improve assessment practice by addressing this aspect more explicitly and constructing assessment standards in communities.

Using the HEA tenets as a lens to investigate the assessment practices on translation modules suggests avenues for further research and development of assessment. The first is to explore the ways in which academics and ‘professional communities set up opportunities and processes’ (HEA, 2012, p. 21) to facilitate the development of professional judgement. Secondly, the authors would like to explore the ways in which assessment literacy would facilitate not only ‘the immediate needs of certification or feedback to students on their current learning, but also contribute in some way to their prospective learning’ (Boud & Falchikov, 2006, p. 400; Boud & Molloy, 2013). The authors believe that students who have a greater assessment literacy and fully understand the criteria and processes involved in assessing translation will be better equipped to apply this knowledge to translations that they or others have com-

Constructing Standards in Communities

pleted. Therefore, developing assessment literacy within translator education can provide the students with the skills necessary for their continuing professional development (CPD), and this knowledge of assessment literacy could also be applied to assessment practices outside the university and provide the knowledge base for mentoring or CPD courses for those already translating professionally.

REFERENCES

- Adab, B. J. (2000). Evaluating translation competence. In C. Schäffner & B. J. Adab (Eds.), *Developing translation competence* (pp. 215–228). Amsterdam: John Benjamins. doi:10.1075/btl.38.20ada
- Assessment Standards Knowledge exchange (ASKe). (n.d.). *Assessment: An ASKe position paper*. Retrieved July 7, 2017, from <https://www.brookes.ac.uk/WorkArea/DownloadAsset.aspx?id=2147552601>
- Beeby, A. (2000). Evaluating the development of translation competence. In C. Schäffner & B. J. Adab (Eds.), *Developing translation competence* (pp. 185–198). Amsterdam: John Benjamins. doi:10.1075/btl.38.18bee
- Boud, D., & Associates. (2010). *Assessment 2020: Seven propositions for assessment reform in higher education*. Australian Learning and Teaching Council. University of Technology, Sidney. Retrieved July 7, 2017, from https://www.uts.edu.au/sites/default/files/Assessment-2020_propositions_final.pdf
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31(4), 399–413. doi:10.1080/02602930600679050
- Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education*, 38(6), 698–712. doi:10.1080/02602938.2012.691462
- Caminade, M., & Pym, A. (1998). Translator training institutions. In M. Baker (Ed.), *Routledge encyclopaedia of Translation Studies* (pp. 280–285). London: Routledge.
- de Groot, A. M. B. (2000). Complex-skill approach to translation and interpreting. In S. Tirkkonen-Condit & R. Jääskeläinen (Eds.), *Tapping and mapping the process of translation and interpreting* (pp. 53–68). Amsterdam: John Benjamins. doi:10.1075/btl.37.06gro
- Elkington, S. (2016). *HEA transforming assessment in higher education summit 2016: Final report*. Retrieved July 20, 2017, from https://www.heacademy.ac.uk/system/files/downloads/hea_assessment_summit_2016_0.pdf
- EMT Expert Group. (2009). *Competences for professional translators, experts in multilingual and multimedia communication*. Retrieved June 13, 2017, from https://ec.europa.eu/info/education/european-masters-translation-emt/european-masters-translation-emt-explained_en
- EMT Expert Group. (2013). *The EMT translator trainer profile competences of the trainer in translation*. Retrieved June 13, 2017, from <http://docplayer.net/14445272-The-emt-translator-trainer-profile-competences-of-the-trainer-in-translation.html>
- Galán-Mañas, A., & Hurtado Albir, A. (2015). Competence assessment procedures in translator training. *The Interpreter and Translator Trainer*, 9(1), 63–82. doi:10.1080/1750399X.2015.1010358

González Davies, M. (2004). *Multiple voices in the translation classroom: Activities, tasks and projects*. Amsterdam: John Benjamins. doi:10.1075/btl.54

Higher Education Academy (HEA). (2012). *A marked improvement: Transforming assessment in Higher Education*. Retrieved June 13, 2017, from https://www.heacademy.ac.uk/system/files/a_marked_improvement.pdf

Higher Education Funding Council for England (HEFCE). (2016). *National Student Survey Results 2016*. Retrieved June 13, 2017, from <http://www.hefce.ac.uk/lt/nss/results/2016/>

Holroyd, C. (2000). Are assessors professionals? *Active Learning in Higher Education*, 1(1), 28–44. doi:10.1177/1469787400001001003

Huertas Barros, E. (2013). *La competencia interpersonal en la formación de traductores en España: Un estudio empírico-descriptivo del trabajo colaborativo durante la primera etapa de formación en Traducción e Interpretación* (Doctoral dissertation). Retrieved December 28, 2017, from <https://hera.ugr.es/tesisugr/22074508.pdf>

Huertas Barros, E., & Buendía Castro, M. (2018). Analysing phraseological units in legal translation: Evaluation of translation errors for the English-Spanish language pair. In S. Gozdz-Roszkowski & G. Pontrandolfo (Eds.), *Phraseology in legal and institutional settings: A corpus-based interdisciplinary perspective* (pp. 41–60). London: Routledge.

Huertas Barros, E., & Vine, J. (2015). Assessing the situation: Closing the gap in translation assessment. *The Linguist*, 54(4), 22–24. Retrieved December 28, 2017, from <http://thelinguist.uberflip.com/i/552383-the-linguist-54-4/21>

Huertas Barros, E., & Vine, J. (2016). Translator trainers' perceptions of assessment: An empirical study. In M. Thelen, G.-W. van Egdom, D. Verbeeck, Ł. Bogucki, & B. Lewandowska-Tomaszczyk (Eds.), *Translation and meaning: New Series*, 41 (pp. 29–39). Frankfurt am Main: Peter Lang.

Huertas Barros, E., & Vine, J. (2018). Current trends on MA translation courses in the UK: Changing assessment practices on core translation modules. *The Interpreter and Translator Trainer*, 12(1). Retrieved from <http://www.tandfonline.com/doi/full/10.1080/1750399X.2017.1400365>

Hurtado Albir, A. (1999/2003). *Enseñar a traducir: Metodología en la formación de traductores e intérpretes*. Madrid: Edelsa.

Hurtado Albir, A. (2007). Competence-based curriculum design for training translators. *The Interpreter and Translator Trainer*, 1(2), 163–195. doi:10.1080/1750399X.2007.10798757

Hurtado Albir, A. (2015a). *Aprender a traducir del francés al español (guía didáctica): Competencias y tareas para la iniciación a la traducción*. Castelló de la Plana: Universidad Jaume I. Servicio De Comunicación y Publicaciones.

Hurtado Albir, A. (2015b). The acquisition of translation competence: Competences, tasks, and assessment in translator training. *Meta: Journal des Traducteurs / Meta: Translators' Journal*, 60(2), 256–280.

Kelly, D. (2002). Un modelo de competencia traductora: Bases para el diseño curricular. *Puentes*, 1, 9–20.

Constructing Standards in Communities

Kelly, D. (2005). *A Handbook for translator trainers: Translation practices explained*. Manchester, UK: St. Jerome.

Kelly, D. (2007). Translator competence contextualized: Translator training in the framework of higher education reform: in search of alignment in curricular design. In D. Kenny & K. Ryou (Eds.), *Across boundaries: International perspectives on translation studies* (pp. 128–142). Cambridge, UK: Cambridge Scholars Publishing.

Kiraly, D. (2013). Towards a view of translation competence as an emergent phenomenon: Thinking outside the box(es) in translator education. In D. Kiraly, S. Hansen-Schirra, & K. Maksymski (Eds.), *New prospects and perspectives for educating language mediators* (pp. 197–224). Tübingen: Gunter Narr Verlag.

Klimkowski, K., & Klimkowska, K. (2012). Towards empowerment in translator education: Students' opinions and expectations of a translation training course. In S. Hubscher-Davidson & M. Borodo (Eds.), *Global trends in translator and interpreter training: Mediation and Culture* (pp. 180–194). London: Continuum.

Lisaitė, D., Vandepitte, S., Maylath, B., Mousten, B., Valdez, S., Castel-Branco, M., & Minacori, P. (2016). Negotiating meaning at a distance: Peer review in electronic learning translation environments. In M. Thelen, G. van Egdom, D. Verbeeck, Ł. Bogucki, & B. Lewandowska-Tomaszczyk (Eds.), *Translation and meaning: New series, 41(1)* (pp. 99–113). Frankfurt am Main: Peter Lang.

Martínez Melis, N., & Hurtado Albir, A. (2001). Assessment in translation studies: Research needs. *Meta: Journal des Traducteurs / Meta: Translators' Journal*, 46(2), 272–287.

Massey, G., & Ehrensberger-Dow, M. (2012). Evaluating the process: Implications for curriculum development. In L. Zybatow, A. Petrova, & M. Ustaszewski (Eds.), *Translationswissenschaft interdisziplinär: Fragen der Theorie und der Didaktik* (pp. 95–100). Frankfurt am Main: Peter Lang.

Massey, G., & Ehrensberger-Dow, M. (2013). Evaluating translation processes: Opportunities and challenges. In D. Kiraly, S. Hansen-Schirra, & K. Maksymski (Eds.), *New prospects and perspectives for educating language mediators* (pp. 157–180). Tübingen: Gunter Narr.

McAlester, G. (2000). The evaluation of translation into a foreign language. In C. Schäffner & B. J. Adab (Eds.), *Developing translation competence* (pp. 229–242). Amsterdam: John Benjamins. doi:10.1075/btl.38.21mca

Nonaka, I. (1991, November). The knowledge-creating company. *Harvard Business Review*, 96–104.

Orlando, M. (2011). Evaluation of Translations in the Training of Professional Translators: At the Crossroads between theoretical, professional and pedagogical practices. *The Interpreter and Translator Trainer*, 5(2), 293–308.

Orlando, M. (2012). Training of professional translators in Australia: Process-oriented and product oriented evaluation approaches. In S. Hubscher-Davidson & M. Borodo (Eds.), *Global trends in translator and interpreter training: Mediation and culture* (pp. 197–216). London: Continuum.

- PACTE. (2003). Building a translation competence model. In F. Alves (Ed.), *Triangulating translation: Perspectives in process-oriented research* (pp. 43–66). Amsterdam: John Benjamins.
- PACTE. (2005). Primeros resultados de un experimento sobre la competencia traductora. In M. L. Romana (Ed.), *AIETI formación, investigación y profesión: II congreso internacional AIETI, Información y documentación* (pp. 573–587). Madrid: AETI.
- Paran, A. (2017). ‘Only Connect’: Researchers and teachers in dialogue. *ELT Journal*, 71(4), 499–508. doi:10.1093/elt/ccx033
- Pavani, S. (2016). *La evaluación sumativa en la didáctica de la traducción: Un estudio empírico en la enseñanza de la traducción español-italiano* (Doctoral dissertation). Retrieved December 28, 2017, from https://ddd.uab.cat/pub/tesis/2016/hdl_10803_400016/stpa1de1.pdf
- Pedrola, M. (1999). Interview with Peter Newmark. In G. Anderman & M. Rogers (Eds.), *Word, text, translation: Liber Amicorum for Peter Newmark* (pp. 17–25). Bristol, UK: Multilingual Matters.
- Price, M., Rust, C., O’Donovan, B., Hindley, K., & Bryant, R. (2012). *Assessment literacy: The foundation for improving student learning*. Oxford, UK: Oxford Centre for Staff and Learning Development.
- Prieto Ramos, F. (2014). Parameters for problem-solving in legal translation: Implications for legal lexicography and institutional terminology management. In *The Ashgate handbook of legal translation* (pp. 135–152). Farnham, UK: Ashgate.
- Rust, C., Price, M., & O’Donovan, B. (2003). Improving students’ learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education*, 28(2), 147–164. doi:10.1080/02602930301671
- Saldanha, G., & O’Brien, S. (2013). *Research methodologies in translation studies*. London: Routledge.
- Schäffner, C., & Adab, B. J. (Eds.). (2000). *Developing translation competence*. Amsterdam: John Benjamins. doi:10.1075/btl.38
- Smith, C. D., Worsfold, K., Davies, L., Fisher, R., & McPhail, R. (2013). Assessment literacy and student learning: The case for explicitly developing students ‘assessment literacy’. *Assessment & Evaluation in Higher Education*, 38(1), 44–60. doi:10.1080/02602938.2011.598636
- Tsagari, D., & van Deemter, R. (Eds.). (2013). *Assessment issues in language translation and interpreting*. Frankfurt am Main: Peter Lang. doi:10.3726/978-3-653-02510-1
- Way, C. (2008). Systematic assessment of translator competence: in search of Achilles’ heel. In J. Kearns (Ed.), *Translator and interpreter training: Issues, methods and debates* (pp. 88–103). London: Continuum International.
- Way, C. (2009). Bringing professional practices into translation classrooms. In I. Kemble (Ed.), *The changing face of translation: Proceedings of the eighth annual Portsmouth translation conference* (pp. 131–142). Portsmouth, UK: University of Portsmouth.
- Way, C. (2014). Structuring a legal translation course: A framework for decision-making in legal translator training. In L. Cheng, K. K. Sin, & A. Wagner (Eds.), *The Ashgate handbook of legal translation* (pp. 135–152). Farnham, UK: Ashgate.

ENDNOTES

- ¹ The terms ‘teacher’ and ‘trainer’ have only been used when they are part of an original quotation. In all other instances the authors have used the term ‘tutor’ to emphasise the teaching role, and ‘educator’ to stress the pedagogical role of tutors.
- ² PACTE’s research project on ‘Establishing Competence Levels in the Acquisition of Translation Competence in Written Translation (NACT)’ is currently developing a proposal of competence descriptors by level.
- ³ Strategic competence (i.e. organisational and planning skills, problem identification and problem-solving, monitoring, self-assessment and revision) (Kelly, 2005: 33).
- ⁴ While Table 1 displays eight competences, there are only seven positions in the ranking since tutors rated both ‘Information mining’ and ‘Strategic competence’ as third in order of importance, and students rated both ‘Information mining’ and ‘Thematic competence’ in the same position (i.e. 4th).
- ⁵ Despite some occasional language issues, the authors have decided to present students’ quotes as originally written by them.
- ⁶ A working platform for submission of coursework where tutors can provide students with personalised feedback and assess their progress throughout the course.

Chapter 12

Introducing a Student Self-Evaluation Grid for Translation Assignments

Mari J. Pakkala-Weckström
University of Helsinki, Finland

ABSTRACT

This chapter will report the results of a study introducing a student self-evaluation grid for translation assignments, based on previous work by Marc Orlando. The grid described here was developed with and for second-year students of English Translation at the University of Helsinki during the autumn terms of 2015 and 2016. This process and the results are described in the light of a pilot study conducted with the students. Based on student feedback, the grid seems to provide a structured framework for evaluating both one's translation process and the translation product, but there are also areas to be developed in this system.

INTRODUCTION

This chapter will report the results of a study introducing a student self-evaluation grid for translation assignments. In the English Translation programme at the University of Helsinki, the students' overall performance in translation courses is assessed using a system based on teacher feedback, peer evaluation, and self-evaluation both during and at the end of each English into Finnish translation course (see Eskelinen & Pakkala-Weckström, 2016 for a detailed description of the system). All in all, this system seems to have worked reasonably well from the point of view of both students and instructors, but nonetheless, the self-evaluation dimension could benefit from some improvement. The method used to enhance student self-evaluation has been a translation commentary for each assignment (see e.g. Fox, 2002; García Álvarez, 2007; Orlando, 2011), combined with a portfolio including a reflective end-of-course self-evaluation (see e.g. Galán Mañas, 2016; Johnson, 2003; Kelly, 2005; Linnankylä, 2001).

Nevertheless, there seems to be a demand for a more focused means for evaluating one's own performance. For example, for some students writing a free commentary after each translation seems to be a

DOI: 10.4018/978-1-5225-5225-3.ch012

Introducing a Student Self-Evaluation Grid for Translation Assignments

burden; they may find it difficult to retrospectively analyse their translation processes (see Eskelinen & Pakkala-Weckström, 2016, p. 327). Furthermore, a translation commentary tends to concentrate on the translation process, and does not necessarily include the evaluation of the product, i.e. the translation itself. Ideally, in order to develop both their translating and overall studying skills, the students should be able to evaluate both, which is what the self-evaluation grid introduced in this chapter aims at. Thus, the main motivation behind this current project has been to assist the students in their self-evaluation by giving them a structured frame of reference to discover the strengths and weaknesses of both their processes and products.

However, while the emphasis has been on raising the students' awareness, the pilot study presented here also clearly suggested that this kind of self-evaluation system designed for students can also benefit instructors in the assessment process. Grading the students' final portfolios becomes both easier and more transparent when the students themselves have laid the groundwork on their strong and weak points and progress during the course.

SELF-EVALUATION AS A TOOL IN TRANSLATOR TRAINING

Assessment regulates learning; therefore, regardless of the method used, assessment should always be constructively aligned, i.e. based on the intended learning outcomes (e.g. Biggs & Tang, 2011, p. 11; Bloxham, 2015, p. 109; Huertas Barros & Vine, 2018a; Kiraly, 2003). According to Brown, Bull and Pendlebury (1997), "assessment defines what students regard as important, how they spend their time and how they come to see themselves as students and then as graduates" (p. 7). Therefore, from the point of view of instructors, who also tend to be in charge of course and syllabus design, assessment methods are crucial in many ways. Kearney suggests that the following questions be asked when making decisions about assessment in Higher Education:

- Why do we assess?
- For whose benefit do we assess?
- Are these achieved through current practices? (2013, p. 876)

One of the points to be considered is agency, i.e. who assesses (see also Brown 2015, p.110)? Traditionally, all assessment is carried out by instructors, although self- and peer-assessment have also gained ground in the past decades (see e.g. Sluijsmans, Dochy, & Moerkerke, 1999). According to Brown, Bull and Pendlebury (1997), "self-assessment is central to life-long learning and the development of professional competence" (p. 178). Biggs and Tang (2011) also stress the connection of self-assessment with professional life: "Making judgements about whether a performance or a product meets the given criteria is vital for effective professional action in any field" (p. 217). Self-assessment can also teach students autonomy, reflection of their own work, and give them responsibility for their own learning (see Race, 2014, p. 93).

There is a wide range of approaches to assessing translations in translator training (see e.g. Galán-Mañas & Hurtado Albir, 2015; Garant & Garant, 2001; Huertas Barros & Vine, 2018a; Kelly, 2005; Martínez Melis & Hurtado Albir, 2001; Orlando, 2011, 2012). The focus in these studies tends to lie mainly on instructor-centred evaluation; in this study, the idea is to shift some of this agency to the students to benefit both students and instructors.

Introducing a Student Self-Evaluation Grid for Translation Assignments

In translator training, the translation assignments and exercises given to students can be directly aimed at simulating professional life. In other words, translator trainees can start actual hands-on practice from very early on in their studies, even though careful monitoring from instructors is important (see Kelly 2007, pp. 138-139). This, however, also requires that all assessment should be carefully designed to support the professional development of the students. According to Orlando (2011):

In a programme aiming at training future professionals along theoretical and practical lines, students and instructors must all understand that there is not just one way to evaluate translations; that the evaluation objectives and criteria may differ from one assignment to another; that a translation is always evaluated according to certain expectations (e.g. textual for the author and the reader, extra-textual for the agency, the client, the client or the commissioner, both textual and extra-textual for an instructor, etc.). Students must be sensitized to this variety of expectations and situations (pp. 296-297).

In addition to sensitizing students to various expectations and situations, it can be argued that students should also learn to proofread and revise their own texts from early on; this is a skill that every translator should possess to ensure the best possible quality of their work. Stressing the importance of systematic self-evaluation tools from the outset of translator training is a way of guaranteeing that this quality monitoring becomes a routine for the students. Way (2008), who has introduced the ‘Achilles’ Heel’ record sheet for students’ self-assessment, notes that to achieve autonomy in their learning, students should have access to a tool which will help them identify their weaknesses and strengths (p. 93). Way’s record sheet is based on Kelly’s (2005) model of translator competence. For self-assessment, Kelly (2005) herself proposes a questionnaire provided by the instructor (p. 143) which is akin to what the programme at the University of Helsinki uses (see Eskelinen & Pakkala-Weckström, 2016, p. 322).

As mentioned above, the assessment of students’ performance in the translation courses in the programme is built around three dimensions: 1) individual feedback from the course instructor on each translation assignment; 2) peer evaluation in the form of class discussions of assignments; and 3) self-evaluation. The peer-evaluation usually comprises a class discussion of a translation chosen by the instructor from the submitted assignments by the group. The translation is always anonymized and the students are expected to participate actively in the discussion and present their own alternative solutions. Class attendance is important, and may have an effect on the students’ final grades (cf. Eskelinen & Pakkala-Weckström, 2016, p. 322). The self-evaluation is usually carried out in two phases: first, in a translation commentary or diary (see e.g. Orlando 2011, pp. 304-305) accompanying each assignment and, second, at the end of each course, as a retrospective evaluation of one’s own progress during the course. This may also contain an element of self-assessment; i.e. the students may propose a grade for themselves (Eskelinen & Pakkala-Weckström 2016, p. 323).

While the students are provided with some general guidelines for both the translation commentaries and the end-of-course self-evaluation, there is a clear need for a system which would help the students pinpoint both their strengths and weaknesses over a longer time period. First of all, retrospective analyses in general tend to have some inherent weaknesses – for example, as Dimitrova and Tiselius (2009) point out, “what subjects will actually report depends on many factors, above all memory constraints, but also for instance the subjects’ energy and willingness to report what they recall” (p. 114; for a summary on retrospection and process research in Translation Studies, see also Massey & Ehrensberger-Dow 2011, pp. 27-29). Massey and Ehrensberger-Dow (2014) further stress that while various deductive commentaries may provide access to translation processes and promote self-reflection, they are difficult to validate (p.

Introducing a Student Self-Evaluation Grid for Translation Assignments

83). Similarly, there seems to be a certain ambiguity in the students' end-of-course reports. When asked to analyse how they have developed as translators, some students find it difficult to become aware of their progress, and others will describe their development in very general terms. Therefore, the role of self-evaluation in the final assessment of students' performance in the course must be carefully considered and can be seen as a challenge for the system, particularly if students are penalised for inadequate commentaries (Eskelinen & Pakkala-Weckström 2016, p. 327).

The self-evaluation grid presented in this chapter is not meant to replace the commentaries or the end-of-course retrospection, but rather to act as a bridge between the two. The idea is that while commentaries are written to accompany translation assignments, the grid is used after class discussions and instructor feedback to analyse both the product (Were there translation errors? Misunderstandings of the source text? Grammatical errors? Inappropriate use of register?) and the process (Were the strategies used effective? Was the research conducted adequate? Was the proofreading accurate?). This differs from the commentary written directly after completing the assignment in that the students will have the benefit of having discussed the assignment in class and having received detailed feedback from the instructor. A more structured system of self-evaluation would provide the students with more concrete information to draw upon for their final self-evaluations as well as more effective tools for monitoring the quality of their work.

MATERIAL AND METHOD

The evaluation grid discussed in this section is based on Orlando's (2011, 2012) but modified to suit the purpose of self-evaluation, as opposed to evaluation by the instructor as is the case of Orlando's. Orlando (2011) introduces two variations for his evaluation grid: grid A, for a product-oriented evaluation, and grid B, for a process-oriented evaluation (pp. 302-303). It should be noted, however, that even though the latter grid is introduced as process-oriented, it also comprises the evaluation of the product. Both grids systematically address a variety of issues from overall comprehension of the source text and accuracy of the target text to more detailed grammar, spelling and punctuation errors in the target text. Both grids are point-based, i.e. students are given a final mark out of 100 by the instructor. Grid B also comprises a section where translation commentaries (or diaries) are analysed in terms of research, strategies and justifications, and revision (Orlando, 2011, pp. 302-303). Since the assessment system used in the English Translation programme at the University of Helsinki focuses on both products and processes, grid B was the starting point for the current study.

The development and testing of the grid was conducted in co-operation with the students of Professional Translation II (from English into Finnish), a second-year general course which precedes the more specialised translation courses (e.g. law, business, science, and technology). This approach is in consonance with developing the notion of "assessment literacy" and constructing standards together in communities of practice (Huertas Barros and Vine, 2018b). The course assignments mainly consist of source texts that are chosen to prepare students for the specialised courses, i.e. popular science texts from various fields. During the course, the students were also exposed to post-editing a machine-translated text and translation memory use. Most assignments were completed as weekly individual homework and the students were aware that the instructor was planning to write a report on the research at a later stage.

There were 12 participants in the course when the grid was piloted in 2015 and 23 in 2016. The raw material of the study comprised the different versions of the self-evaluation form, examples of the grids

completed by students, teacher feedback and, most importantly, the students' feedback on the evaluation grid which was specifically requested at the end of both courses. The system underwent some significant changes after the pilot study in 2015; the different stages of the process, along with the practical and pedagogical reasons for the changes made to the system shall be outlined in this section. The student feedback will be analysed both quantitatively and qualitatively. It is to be noted that the development of the system is still underway.

In the autumn of 2015, the students were asked about their willingness to participate in the current project. The instructor introduced Orlando's grid B (Table 1) to the group, and its suitability for self-evaluation was discussed in class. It should be noted that self-evaluation does not, in this context, involve grading, but instead simply aims at helping the students be aware of possible weaknesses in their translations. Also, it was emphasized that the self-evaluations would not affect the students' final grades. Since the students were positive about the project, the instructor translated the grid into Finnish, taking into consideration the points that were raised during the class discussion (Table 2).

In order to accommodate the grid to meet the demands of the current project, some initial alterations were made: first, since the grid was aimed at facilitating the students' self-assessment, it seemed only sensible to replace the points by a Likert scale (from one to three, for none [1] – some [2] – several [3]), as penalizing one's own work does not present itself as a motivating form of self-assessment. Furthermore, since the purpose of self-evaluation is not giving grades but rather raising the students' awareness of their own strengths and weaknesses, points and final marks were not relevant for the purpose of the grid. Second, since the students (unlike those of Orlando) were translating into their mother tongue, grammatical or syntactic choices which produce distortion to the meaning are rare, therefore, this particular section in the grid was given less priority, and spelling and punctuation errors were combined. In the section 'overall translation effect' some adjustments were also made: e.g. the aspects 'appropriateness for target audience' and 'adherence to brief' were combined and the terms *Skopos* and *function* were added. Finally, a section for the translator's own retrospective observations on the assignment was added to the end (Table 2).

The first trial of the grid was conducted in connection with the students' fourth homework assignment (introducing the idea of self-evaluation of translations with a specific tool as well as discussing possible modifications to the original grid took some weeks). In addition to using the Likert scale, the students were told that they could use the own observations field for verbal comments on each point, but this was voluntary. After this first trial, and following another class discussion, some further alterations were made. First of all, the scale was altered to comprise five alternatives. The students strongly felt that the three-point scale was not enough; if they had made one or two errors in e.g. punctuation, they would have to choose '2', which was already mid-way on the scale. Also, since no further instructions considering the number of errors were given, they questioned how they could know how many would constitute '3', i.e. 'several'?

Another alteration to the grid was adding some clarifying instructions of what was referred to under the various headings; for example, when students translate into their native language, their grammatical mistakes are usually minor ones, like subject-verb agreement errors. These may often be result of careless revision and, therefore, difficult to notice in one's own text. In the 'Translation Effect' section the students could list 'insertions or omissions which improve the target text' – these could be e.g. explicitations to clarify the text for the target text audience, or omissions to achieve the same effect. These instructions are indicated with italics in the grid (Table 3). This was to add some emphases on the positive aspects of the translations.

Introducing a Student Self-Evaluation Grid for Translation Assignments

Table 1. Grid B: For a process-oriented evaluation

Translation Exercise /50
<ul style="list-style-type: none"> ● Overall Comprehension of Source Text (misinterpretations with more or less effect on accuracy) 0 2 4 6 8 10 ● Overall Translation Accuracy / Transfer ST>TT (mistranslations with more or less effect on accuracy) 0 2 4 6 8 10 ● Omissions/ Insertions (with more or less effect on accuracy) 0 2 4 6 8 ● Terminology / Word Choices (affecting more or less the localized meaning) 0 2 4 6 ● Grammatical Choices / Syntactic Choices (producing more or less distortion to the meaning) 0 2 4 6 ● Spelling Errors 0 2 4 6 ● Punctuation Errors 0 1 2 ● Formatting Errors 0 1 2
Overall Translation Effect /20
<ul style="list-style-type: none"> ● Appropriateness for Target Audience 0 2 4 6 8 ● Readability / Idiomatic Correctness 0 2 4 6 8 Adherence to Brief: ● Function / Completeness 0 1 2 ● Style / Presentation / Genre 0 1 2
Translator's Strategy /30
INTEGRATED TRANSLATOR'S DIARY (Reporting of problems, actions, decisions) <ul style="list-style-type: none"> ● Use of material / research 0 2 4 6 8 10 ● Strategy / Justifications / Report (relevance of choices) 0 2 4 6 8 10 ● Revision / Proofreading 0 2 4 6 8 10
OVERALL COMMENT: FINAL MARK: /100

Orlando 2011, p. 303.

Additionally, some points, such as source text comprehension and translation accuracy were combined to keep the grid manageable. Finally, a point on conversions (for measurements, etc.) was added, since this is an area where Finnish students often seem to have some difficulties.

ANALYSIS

In the following sections the feedback on the self-evaluation grid from 2015 and 2016 are analysed in some detail. The groups are treated separately, as the 2015 students were a pilot group actively involved in the development of the grid. For the 2016 group, the grid was already an established part of the evaluation system, which was explained to them at the start of the course. Both groups were, however, specifically asked to give feedback on the grid in their final portfolios. The feedback extracted from these portfolios

Introducing a Student Self-Evaluation Grid for Translation Assignments

Table 2. Adapted evaluation grid, version 1

Translation Evaluation Grid/MPW
Translation
<ul style="list-style-type: none"> ● Comprehension of source text (misinterpretations?) none 1 2 3 several ● Translation accuracy (mistranslations?) none 1 2 3 several ● Omissions / insertions (which affect understanding?) none 1 2 3 several ● Lexical problems: terms/ word choices / proper names, etc. none 1 2 3 several ● Grammatical errors (which distort the meaning?) none 1 2 3 several ● Spelling and punctuation errors none 1 2 3 several
Translation Effect
<ul style="list-style-type: none"> ● Adherence to brief and taking the intended target audience into account (skopos / function) taken well into account 1 2 3 not taken into account ● Readability and idiomatic correctness very good 1 2 3 not good ● Style, genre taken well into account 1 2 3 not taken into account
Translator's Strategies (From the Translation Commentary)
<ul style="list-style-type: none"> ● Research very good 1 2 3 not good ● Strategies and justifications very good 1 2 3 not good ● Proofreading very good 1 2 3 not good
Own observations

Translated from Finnish by the author.

Source: Orlando, M. 2011. Evaluation of Translations in the Training of Professional Translators. At the Crossroads between Theoretical, Professional and Pedagogical Practices. *The Interpreter and Translator Trainer* 5(2). 293-308.

is analysed both quantitatively and qualitatively below. Both sets of students have been given individual numbers randomly in order to preserve their anonymity. The portfolios were submitted through Moodle, and their content is visible only to the instructor and each student individually. All citations concerning the evaluation grid from the reports have been translated from Finnish by the author and have been used to support the analysis and discussion of the results.

The feedback from the students has been attributed to one of three possible categories: positive, neutral and negative. It should be noted that when dividing the feedback into these categories, the author has relied on her own interpretation, so the categorization is undoubtedly subjective and, in particular, some of the comments categorized as 'neutral' could possibly have been placed in one of the other categories.

Data From 2015

The 2015 group comprised 12 students. While all were native Finnish speakers, the group was rather heterogenous in terms of educational and professional background and age. All twelve students completed the course successfully. It should be noted that the group size was smaller than average (the annual intake of students of English Translation at the University of Helsinki is around 20).

Introducing a Student Self-Evaluation Grid for Translation Assignments

Table 3. Adapted evaluation grid, version 2

Translation Evaluation Grid/MPW
Translation
<ul style="list-style-type: none"> ● Comprehension of source text (misinterpretation?)/ Translation accuracy (mistranslations?) none 1 2 3 4 5 several ● Omissions / insertions (which affect understanding?) none 1 2 3 4 5 several ● Lexical problems: terms/ word choices / proper names etc. <i>Geographical names, historical persons</i> none 1 2 3 4 5 several ● Grammatical errors (which distort the meaning?) <i>inflections, pronouns, agreement</i> none 1 2 3 4 5 several ● Conversions <i>measurements, conventions</i> none 1 2 3 4 5 several ● Spelling and punctuation errors none 1 2 3 4 5 several
Translation effect
<ul style="list-style-type: none"> ● Adherence to brief and taking the intended target audience into account (skopos / function) <i>insertions / omissions which improve the target text</i> <i>text conventions</i> <i>domestication (localization) / foreignization</i> Readability and idiomatic correctness <i>metaphors, word-play</i> ● Style, genre, register <i>collocations</i>
Translator's Strategies (From the Translation Commentary)
<ul style="list-style-type: none"> ● Research very good 1 2 3 not good ● Strategies and justifications very good 1 2 3 not good
Own observations

Translated from Finnish by the author.

Source: Orlando, M. 2011. Evaluation of Translations in the Training of Professional Translators. At the Crossroads between Theoretical, Professional and Pedagogical Practices. *The Interpreter and Translator Trainer*, 5(2). 293-308.

Even though the students were instructed – both orally in class on several occasions, and in writing in the portfolio instructions in Moodle – to give feedback on the self-evaluation grid in their portfolios, one student failed to do so. This is the reason why eleven reports on the reception of the grid are analysed here. The breakdown of the categorisation of the students' comments was as follows:

The category 'positive' includes those three comments which portrayed an overall positive attitude towards the self-evaluation grid; however, suggestions for improvement were also given by two students (2015/3 and 2015/12), and only one can be said to be clearly positive without reservations:

(1) The self-evaluation grid was a good addition to the teaching – it made us focus more deeply on our translations and areas to be developed (2015/10).

The 'neutral' category is somewhat mixed; it includes one student (2015/1) who criticized the first version, but was cautiously optimistic about the second version:

Introducing a Student Self-Evaluation Grid for Translation Assignments

Table 4. The reception of the self-evaluation grid in 2015

POSITIVE	3
NEUTRAL	4
NEGATIVE	4
TOTAL	11

(2) *When I was filling in the grids at the end of the course, after class discussions and comments, it was easier to do and I was better able to count and analyse my mistakes. I noticed a bit more systematically what kind of mistakes I make, and what I should pay more attention to.*

The other three students had various reservations; student 2015/5 considered the grid 'OK', but clearly preferred the 'old' system of free commentaries; student 2015/7 could see that the system based on Likert scales might benefit others, but did not include her/himself in that group. Student 2015/8 had reservations about self-evaluation, but also saw some benefits:

(3) *The grid is, however, quite a good tool for evaluating a translation as a whole; even though no one probably notices all their mistakes, filling in the grid provides some sort of framework.*

The negative comments included aversion to self-evaluation in general, and also considered the grid additional work. One student (2015/6) was concerned that the self-evaluation might have an effect on the instructor's opinion of the translation, and would have preferred to use the grid for blind peer evaluation. It needs noting, however, that at least one of the students who gave negative feedback seemed to have misunderstood the instructions, which encouraged the student to add comments below the scales, despite not being necessary to do so:

(4) *I find it difficult to evaluate my performance by numerical scales, and filling in the grids did not make me approach my translation from a new angle. I might have preferred answering the questions verbally; then I would have had to analyse my translation and working methods more carefully (2015/4).*

All in all, the students gave valuable suggestions – both direct and implied in their critical comments – for improving the system. Student 2015/3 felt that the grid was useful, but also made the following observation:

(5) *The style of the grid was without question somewhat negative, listing only so-called 'errors' [...] in the future, the positive aspects of translations should somehow be included in the grid if possible.*

Student 2015/12 also noted that using the grid was extremely time-consuming, and gave concrete suggestions for improving the system (it should be noted that the student in question perhaps took the self-evaluation more seriously than the other students in the group, listing errors and commenting on every section of the grids in detail):

Introducing a Student Self-Evaluation Grid for Translation Assignments

(6) Evaluating all texts took A LOT of time and was sometimes excruciating. So, thinking of future students, it would be more humane (...) if the amount of compulsory self-evaluations was halved, or alternatively more credits were given for the course.

When the students were initially introduced to the self-evaluation grid, they were instructed to fill it in immediately after they had completed their translation assignments and written their translation commentaries (see also above). This, however, seems to have been confusing for some students as is demonstrated by the following feedback from two students:

(7) I personally don't see much use in filling in the grids immediately after translating. As I commented when first using the grid, it is impossible to find (at least all!) grammar mistakes, since of course I have made my translation as fluent as possible. So, I am 'blind' to my mistakes (2015/1).

(8) I personally disliked the self-evaluations: if done immediately after translating, I would evaluate my [translation] too positively (after all, I have just done "the best possible translation" I can do) (2015/6).

The early class discussions after the introduction of the grid brought up some confusion – the students felt that trying to spot errors in a translation they had just completed according to their best abilities was something of a futile exercise. This concern is also expressed in examples 7 and 8 and, as a result, the system was modified accordingly mid-course; the students were instructed to fill in the grid after class discussion and after having received individual feedback from the instructor. This seemed a more prudent approach; however, one student (2015/6) did not find this practice particularly useful either, noting that she/he would just repeat the instructor's comments in the self-evaluation.

Overall, the data from 2015 suggests the self-evaluation grid was received with mixed feelings. Based on the students' experiences and feedback, it was apparent that it would be more beneficial for the students to perform their self-evaluation after class discussions and the instructor's feedback. The workload, however, should also be taken into account; filling in the grid for several translations after class discussion can be time-consuming. However, the students also appeared to actually benefit from a more systematic approach to analysing their translations, so bearing their valuable feedback in mind, the author decided to continue with the experiment with the 2016 course.

Data From 2016

The data from autumn 2016 comprises portfolios by 23 students. Twenty students were second- or third-year English Translation majors, and three were new students in the MA Programme for Translation Studies and Multilingual Communication¹. The students submitted portfolios containing a selection of three freely chosen translations in two versions (original and corrected), the self-evaluation grids for each three, and the end-of-course reflective commentary (see previous section). All 23 students completed the course successfully. As with the previous year's group, the students were encouraged – both orally and in writing in instructions on the course Moodle site – to give feedback on the self-evaluation grid in their portfolios, but two students nevertheless failed to do so; therefore, the 2016 data presented comes from 21 student portfolios. The breakdown of the categorisation of students' comments (i.e. positive, neutral, negative) was as follows:

Introducing a Student Self-Evaluation Grid for Translation Assignments

Table 5. The reception of the self-evaluation grid in 2016

POSITIVE	14
NEUTRAL	4
NEGATIVE	3
TOTAL	21

Altogether 14 students gave feedback categorized as ‘positive’. However, only four out of these 14 can be called ‘positive without any reservations’, as exemplified by the following comments:

(9) The separate section headings made spotting mistakes easier than by normal proofreading. One usually becomes blind to one’s own mistakes when proofreading, but when going through the text looking for spelling mistakes, for example, it is easier to find problem spots. Assessing mistakes will certainly also improve the final quality of the translation (2016/9).

(10) The self-evaluation grid was, in my opinion, good, clear, and useful. I would be happy to have such a tool in future to evaluate my translations (2016/13).

(11) Using the self-evaluation grid was easy, I did not have any problems with it. The grid covered everything that should be taken into account in the translation in terms of grammar and interpretation (2016/14).

(12) The self-evaluation grid was useful for the portfolio, as it makes writing the translation commentaries more organized, and focuses on one aspect at a time. It was nice to edit the texts written during the course, and get a new angle on the texts. The self-evaluation grid helped to pay closer attention to the text, and I was glad to be able to edit the texts so that they turned out clearer and more concise (2016/15).

In the rest of the feedback classified as ‘positive’, there were reservations or suggestions for improvement, even though the overall impression of the comments was still positive, albeit for various reasons. For example, one student clearly had an aversion towards self-evaluation, but found the grid useful for that very reason:

(13) When it comes to the self-evaluation grid... In my opinion, I am very bad at self-evaluations, and I do not like them. My philosophy is that my opinion does not really count, the most important thing is to convince my future employer that my translations are brilliant. Therefore, I also feel that any tools that will make self-evaluation easier are most welcome (2016/2).

In the following comments, the students appear to find the grid useful, but are somewhat critical about self-evaluation through the Likert scale:

(14) The self-evaluation grid was a new thing for me. I think it is good that the evaluation of a translation is divided into different categories, because there are so many aspects to consider (grammar, structure,

Introducing a Student Self-Evaluation Grid for Translation Assignments

metaphors etc.). I feel, however, that for me personally, evaluating my work on a numerical scale is somewhat difficult (2016/14).

(15) In my opinion, the self-evaluation grid needs more sections where you can evaluate yourself in other ways, and not numerically, and the evaluation in the last section could be more profound, so that one would really have to elaborate on one's performance. The current one does the job, but is, in my opinion, a bit rigorous (2016/5).

(16) The self-evaluation grid has been useful in analysing my translations, but giving numerical assessments of myself has been difficult (2016/11).

Two students in the 'positive' category pointed out that the grid was additional work, although limiting its use to just three translations seemed to meet with their approval. The aspect of extra work is also implied in the case of the student who suggested improvements to the grid to make it shorter.

(17) Filling in the grid was also quite laborious, so it is good that it is only required for the translations chosen for one's final portfolio (2016/10).

(18) The self-evaluation grid helped to read the texts closely, and analyse what actually caused the problems with the translation. Three translations were enough, since filling in [the grids] was quite laborious (2016/17).

(19) The grid could perhaps be developed to combine similar sections, thus making it shorter. For example, questions on grammar mistakes and spelling errors could be combined. Now it felt like you were giving the same answers to several questions (2016/11).

Four students gave feedback which was classified as 'neutral', i.e. it was neither positive nor negative overall. One student found several things to criticize, but did not altogether reject the idea behind the system:

(20) The self-evaluation grid was more confusing than helpful. I have never been able to take seriously forms which require one to evaluate one's performance numerically, since how can that be done objectively? I also felt that three main sections in the open questions (translation effect, readability and idiomatic expression, genre/ style/ register) overlapped so that I wrote the same things in many places. I also wasn't sure whether one needed to write comments under the sub-headings or freely, so I only commented under the sub-headings. All in all, the self-evaluation grid is a good idea, but needs developing in formulating the questions and giving instructions (2016/1).

Student 2016/3 found both positive and negative aspects in using the grid (see example 23), but ended the feedback by stating that the free commentary is better. Student 2016/18 gave feedback which mainly expressed some confusion about the grid, but left it unclear whether the student found the system useful or not. One student listed several negative aspects – for example, numerical evaluation, lack of clarity, extra work – which is why the feedback has been categorized as 'neutral'; however, this particular

Introducing a Student Self-Evaluation Grid for Translation Assignments

comment on the positive aspects demonstrates an ability to think forward towards professional life (cf. Brown, Bull, & Pendlebury, 1997, pp. 178-179):

(21) On the positive side, the grid enabled me to practise analysing my translation using professional terms. When filling in the grid, I noticed that sometimes I lacked the relevant terminology when describing problem areas. In professional life, commenting and justifying one's translation and translation solutions in a credible way is important, which is why it is good to practise these skills already while studying. When filling in the self-evaluation grid one was forced to pay attention to one's problem areas in detail (2016/12).

Of the three students whose feedback was classified as 'negative', one (2016/8) felt that the grid could have been used for blind peer evaluation, another (2016/19) felt the grid required considerable amount of work, and also showed a general aversion towards self-evaluation. The third negative feedback was exceptionally detailed, and contained several suggestions for improvement:

(22) These self-evaluation grids were, in my opinion, rather exhausting, maybe because I had already written translation comments on each text. It did not seem to add anything to this. The grid also seems to focus on actual errors, while successful solutions are left to the very end. The successes could have been dealt with in the same way as the failures, for example in the "omissions / insertions (with more or less an effect on accuracy)". Also, in the case of "research" the scale 1 to 3 is too narrow; I felt I always had to choose '2', since calling you own research 'excellent' would have been odd – surely there is always room for improvement. So, the scale 1 to 5 used elsewhere would have worked better here as well, I think. There were also some unclear things, like "Translator's Strategy", I was not exactly sure whether I was supposed to fill that in myself. I assumed so, but it wasn't made clear. Additionally, I can't really see how I could evaluate my strategies. Perhaps a line or two of instruction could be added here to help us simpletons? (2016/16).

From the data collected regarding the 2016 course, it can be said that the reception of the self-evaluation system was cautiously positive, and the students' input – regardless of whether their feedback fell into the 'positive', 'neutral' or even 'negative' categories – was in many ways constructive and helpful for the instructor. Altogether, seven students considered the self-evaluation grid to be extra work, and five students brought up the issue of evaluation by numerical scales, which they criticized for various reasons: four of them simply found evaluating their own performance on the Likert scale to be lacking objectivity, difficult, or uncomfortable. In example 22, the student only criticizes the three-point scale used in the section concerning research. Other critical observations included unclear instructions, a certain amount of repetition or overlapping (especially concerning the "Translation Effect" section), and excessive emphasis on the negative aspects in the translations. Therefore, a sustained effort and a close collaboration between students and the instructor is essential for self-evaluation and portfolios to be successful (Galán-Mañas, 2016, pp. 161).

On the other hand, several students explicitly stated that they found the system to be clear and helpful (examples 9-12 and 14). Linking self-evaluation to professional life (Biggs & Tang, 2011; Brown, Bull, & Pendlebury, 1997, p. 178) and also the development as a translator came up directly in example 21, and implicitly in the following examples:

Introducing a Student Self-Evaluation Grid for Translation Assignments

(23) I think the self-evaluation grid is a good tool in that it gives guidelines as to what should be looked at when evaluating a text (2016/3).

(24) I don't think the self-evaluation grid is actually a bad idea. It makes the student evaluate the relevant aspects of a translation, which might be missed in free self-evaluation or a translation commentary (2016/7).

(25) The self-evaluation grid was a good tool for evaluating translations. Self-evaluation helped to analyse one's translation strategies, and how many and what kind of mistakes one is prone to make (2016/10).

(26) All in all, the grid helps keep translation strategies in my mind, and encourages me to analyse my solutions: a translation is not simply done, instead, the process involves taking note of challenges and making carefully considered decisions about meeting them. I mainly saw the grid as a useful tool that 'forces' me to revisit my translation and analyse the choices I had made (2016/11).

These comments would seem to indicate that the students understand the translation process as something that can and should be monitored and developed, as suggested by e.g. Galán-Mañas, (2016); and, on the other hand, they also see the translation product as something that has a quality that can and should be evaluated (see also examples 9 and 12 above).

CONCLUSION

This chapter has reported the results of pilot study of a self-evaluation grid for student translations (based on Orlando, 2011, 2012). Two different versions of the grid were tested with the help of mainly second-year English Translation students at the University of Helsinki in 2015 and 2016. The aim was to provide the students with a useful tool for self-monitoring as well as quality control; a grid which will help them conduct detailed retrospective analysis of their strengths and weaknesses in translating, as has proven to be successful in translator training (e.g. Way, 2008). The results can be said to be overall positive, since several students found the tool helpful. However, there is still obviously ground to cover. For example, the students' self-evaluations will always be subjective to some degree, so they will need to be treated as such by the instructor and will require careful monitoring as suggested by Kelly (2007, pp. 138-139). Also, while some students will see the benefits of the system, others may consider it a burden.

Comparing the student feedback from 2015 and 2016, it appears that the development of the self-evaluation system had been going in the right direction; while the 2015 group gave mostly neutral or negative feedback, the data from 2016 shows that the reception of the grid was mainly positive, even though the students also gave a substantial amount of constructive criticism. The obvious difference in the general attitude towards the system can perhaps be explained by the fact that the first set of students actively participated in its development while for the second group the grid was a given feature (although they too were aware that the system was still under development). The 2015 students were invaluable in pointing out weaknesses in both the grid itself as well as initial instructions for its use; the second set of students had a clear advantage considering their starting point.

Still, the results of the current project are optimistic, since several students considered the self-evaluation grid useful in detecting development areas in their translations and translation processes. It is important that translator trainees should learn from early on to systematically evaluate their own performance – both product and process – and this kind of structured evaluation method would appear to provide a useful tool for this kind of evaluation.

However, the student feedback received from both the 2015 and 2016 groups also indicates that there is still room for improvement in the system. According to the feedback, there are certain minor modifications that should be done: first of all, the instructions for using the grid should obviously be clearer, and they should perhaps be repeated more often in class. Second, the positive aspects of translations could perhaps be accentuated even more in the grid, although the second version contains an improvement in this respect.

A more fundamental challenge in the system is whether to preserve the Likert scale, since several students reported an aversion towards numerical self-evaluation. The issue obviously needs to be addressed; perhaps it could simply be remedied by clearer instructions, for example by explicating the number of errors that correspond to a point on the scale. Also, the students' concern about the lack of objectivity is perhaps exaggerated since their self-evaluation grids do not have an effect on their final grades; stressing this could also relieve this particular anxiety. Still, one needs to consider carefully the issue whether 'grading' one's mistakes is actually relevant? If not, how does one still include the element of a systematic and structured frame of reference?

Also, since the feedback also indicated that many students considered the grid additional work, it would be important to find a balance – not to overburden students, but aid them in their development towards professionalism and the best possible quality in their work.

Finally, it must be noted that although the results of this study are promising, and indicate that developing the self-evaluation grid further would be beneficial to future students, the data is limited, and because of the small number of students available for the experiment, it was not possible to include a control group; nor were the students' comments cross-referenced with their final grades. It should be further noted that since their final self-evaluation has an effect on the grades, at least some of the students may have given more positive feedback on the grid than they would otherwise have done. These reservations should be taken into account when developing the system further.

REFERENCES

- Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). Berkshire, UK: Open University Press.
- Bloxham, S. (2015). Assessing assessment. New developments in assessment design, feedback practices and marking in higher education. In H. Fry, S. Ketteridge, & S. Marshall (Eds.), *A handbook for teaching and learning in higher education: Enhancing academic practice* (4th ed.; pp. 107–122). London: Routledge.
- Brown, G., Bull, J., & Pendlebury, M. (1997). *Assessing student learning in higher education*. London: Routledge.

Introducing a Student Self-Evaluation Grid for Translation Assignments

Brown, S. (2015). *Learning, teaching and assessment in higher education: Global perspectives*. London: MacMillan Education. doi:10.1007/978-1-137-39667-9

Englund Dimitrova, B., & Tiselius, E. (2009). Exploring retrospection as a research method for studying the translation process and the interpreting process. In I. M. Mees, F. Alves & S. Göpferich (Eds.), *Methodology, technology and innovation in translation process research: A tribute to Arnt Lykke Jakobsen* (pp. 109-134). Fredriksberg: Samfundslitteratur.

Eskelinen, J., & Pakkala-Weckström, M. (2016). Assessing translation students' learning of professional competences. *Translation Spaces*, 5(2), 314–331. doi:10.1075/ts.5.2.08esk

Fox, O. (2000). The use of translation diaries in a process-oriented translation teaching methodology. In C. Schäffner & B. Adab (Eds.), *Developing translation competence* (pp. 115–130). Amsterdam: John Benjamins. doi:10.1075/btl.38.12fox

Galán-Mañas, A. (2016). Learning portfolio in translator training: The tool of choice for competence development and assessment. *The Interpreter and Translator Trainer*, 10(2), 161–182. doi:10.1080/1750399X.2015.1103108

Galán-Mañas, A., & Hurtado Albir, A. (2015). Competence assessment procedures in translator training. *The Interpreter and Translator Trainer*, 9(1), 63–82. doi:10.1080/1750399X.2015.1010358

Garant, J., & Garant, M. (2001). Case study: Variation of assessment practices in translator education. In P. Kukkonen & R. Hartama-Heinonen (Eds.), *Mission, vision, strategies, and values: A celebration of translator training and translator studies in Kouvola* (pp. 45–58). Helsinki: Helsinki University Press.

García Álvarez, A. M. (2007). Evaluating students' translation process in specialised translation: Translation commentary. *The Journal of Specialised Translation*, 7, 139–163.

Huertas Barros, E., & Vine, J. (2018a). Current trends on MA translation courses in the UK: Changing assessment practices on core translation modules. *The Interpreter and Translator Trainer*, 12(1). Retrieved from <http://www.tandfonline.com/doi/full/10.1080/1750399X.2017.1400365>

Huertas Barros, E., & Vine, J. (2018b). Constructing standards in communities: Tutors' and students' perceptions of assessment practices on an MA translation course. In E. Huertas Barros, S. Vandepitte & E. Iglesias Fernández (Eds.), *Quality assurance and assessment practices in translation and interpreting: Advances in linguistics and communication studies (ALCS)*. Hershey, PA: IGI Global.

Johnson, J. (2003). Learning through portfolios in the translation classroom. In J. Baer & G. S. Koby (Eds.), *Beyond the ivory tower: Rethinking translation pedagogy* (pp. 97–116). Amsterdam: John Benjamins. doi:10.1075/ata.xii.08joh

Kearney, J. (2013). Improving engagement: The use of 'authentic self and peer-assessment for learning' to enhance the student learning experience. *Assessment & Evaluation in Higher Education*, 38(7), 875–891. doi:10.1080/02602938.2012.751963

Kelly, D. (2005). *A Handbook for translator trainers*. Manchester: St. Jerome.

- Kelly, D. (2007). Translator competence contextualized. Translator training in the framework of higher education reform: In search of alignment in curricular design. In D. Kenny & K. Ryou (Eds.), *Across boundaries. International perspectives on translation studies* (pp. 128–142). Newcastle: Cambridge Scholars Publishing.
- Kiraly, D. C. (2003). From instruction to collaborative construction: A passing fad or the promise of a paradigm shift in translator education? In B. J. Baer & G. S. Koby (Eds.), *Beyond the ivory tower: Rethinking translation pedagogy* (pp. 3–27). Amsterdam: John Benjamins. doi:10.1075/ata.xii.03kir
- Linnankylä, P. (2001). Portfolio: Integrating writing, learning and assessment. In P. Tynjälä, L. Mason, & K. Lonka (Eds.), *Studies in writing: Writing as a learning tool: Integrating theory and practice* (Vol. 7, pp. 145–160). Dordrecht: Kluwer Academic Publishers. doi:10.1007/978-94-010-0740-5_9
- Martínez Melis, N., & Hurtado Albir, A. (2001). Assessment in translation studies: Research needs. *Meta*, XLVI(2), 272–287. doi:10.7202/003624ar
- Massey, G., & Ehrensberger-Dow, M. (2011). Commenting on translation: Implications for translator training. *The Journal of Specialised Translation*, 16, 26–41.
- Massey, G., & Ehrensberger-Dow, M. (2014). Looking beyond text. The usefulness of translation process data. In D. Knort, C. Heine, & J. Engberg (Eds.), *Methods in writing process research* (pp. 81–98). Frankfurt am Main: Peter Lang.
- Orlando, M. (2011). Evaluation of translations in the training of professional translators: At the crossroads between theoretical, professional and pedagogical practices. *The Interpreter and Translator Trainer*, 5(2), 293–308.
- Orlando, M. (2012). Training of professional translators in Australia: Process-oriented and product-oriented evaluation approaches. In S. Hubscher-Davidson & M. Borodo (Eds.), *Global trends in translator and interpreter training* (pp. 197–216). London: Bloomsbury Academic.
- Race, P. (2014). *Making learning happen: A guide for post-compulsory education* (3rd ed.). Los Angeles: Sage.
- Sluijsmans, D., Dochy, F., & Moerkerke, G. (1999). Creating a learning environment by using self-, peer-, and co-assessment. *Learning Environments Research*, 1(3), 293–319. doi:10.1023/A:1009932704458
- University of Helsinki. (2017, July 13). *Line B, BA Degree in a Discipline not Related to Translation*. Retrieved from <http://blogs.helsinki.fi/translationstudies/files/2014/12/structure-of-degree-programme.pdf>
- Way, C. (2008). Systematic assessment of translator competence: In search of Achilles' heel. In J. Kearns (Ed.), *Translator and interpreter training: Issues, methods and debates* (pp. 88–103). London: Continuum.

KEY TERMS AND DEFINITIONS

Portfolio Assessment: The assessment of student performance through a portfolio, which contains a collection of their translations and a retrospective analysis of their performance during the course. The guidelines for the portfolios are discussed with the students at the start of each course.

Self-Evaluation: A student's own evaluation of their performance which does not involve grades.

Student Feedback: A student's evaluation of the course, included in the portfolio.

Translator Training: A university-level program for training professional translators.

ENDNOTE

- ¹ See <http://blogs.helsinki.fi/translationstudies/files/2014/12/structure-of-degree-programme.pdf>
Line B, BA Degree in a Discipline not Related to Translation.

Chapter 13

Assessing Translation Students' Reflective and Autonomous Learning

Ya-Yun Chen
Newcastle University, UK

ABSTRACT

As an essential soft skill in life-long learning, reflective and autonomous learning has been an integral part of many translator training programs today. However, how it could be assessed systematically and what factors might influence its acquisition is still much under-researched. To help bridge this gap, this chapter aims at reporting the findings of an empirical study, which used diary, think-aloud and small-group discussion as reflective learning methods to investigate translation students' reflective learning. It first provides an overview of relevant theory and then reports how students' reflective levels were assessed and analyzed quantitatively and qualitatively. Based on the empirical findings, it discusses the factors influencing the success of a reflection-encouraging learning environment, followed by a provisional model of translation students' reflective and autonomous learning process.

INTRODUCTION

Nowadays, with the professional landscape developing rapidly, translation students must meet the changing requirements of translation markets. To facilitate students' professional development beyond their immediate training setting and in the long run, concepts such as lifelong learning and learner autonomy have been upheld and integrated in translator training through a variety of pedagogical strategies to help students develop problem-solving and other transferrable skills (Washbourne, 2014). Amongst them is reflective thinking, which is regarded as essential for professional development (Moon, 1999).

Over the years, and most notably since the 1990s, with the incorporation of contemporary learning theories, translator training has been through a major pedagogical shift, from teacher-centered, translation-focused approaches to more learner-centered, process-oriented ones (Kelly, 2005). The latter often upholds reflective learning activities, ranging widely from group learning to individual learning.

DOI: 10.4018/978-1-5225-5225-3.ch013

Assessing Translation Students' Reflective and Autonomous Learning

Group-based activities are usually justified on the benefits of learning through peer interaction, such as group discussion and collaborative translation projects (e.g. Chen, 2010; Kiraly, 2000, 2003, 2005). Individual-based learning activities, such as learning diary (Fox, 2000; Orlando, 2011), think-aloud techniques (Wakabayashi, 2003) and portfolio (Galán-Mañas, 2016; Johnson, 2003), usually stress how self-assessment and self-awareness help students improve translation quality and their learning skills. Some scholars have proposed a combination of both group and individual learning activities to accommodate a wider range of learning needs and styles (González Davies, 2004; Kelly, 2005).

However, although reflective learning has been integrated into many translator training programs, there is no sufficient empirical research into students' experience with reflection-promoting approaches. It is also unclear how translation students reflect, how their reflective ability can be assessed, and whether reflective learning is conducive to their translation skills and, as a result, translation quality.

This chapter will report on the results of an empirical study specifically designed to investigate students' perception of three selected reflective learning environments, namely, diary, think-aloud (TA), and small-group discussion (SGD), through both quantitative and qualitative methods aiming at understanding students' reflection, how it can be assessed, and the factors influencing their reflective and autonomous learning.

KEY CONCEPTS

Autonomy

Since the shift in translation pedagogy, it can be said that learner autonomy has become “a key concept” in translator training. Helping students develop autonomous learning skills has been regarded as essential for students to conduct more complex translation projects (Kelly, 2005), and, ultimately, ensure that they have the ability to continue to learn after leaving the training program (Kiraly, 2000).

One frequently cited definition of autonomy is perhaps the one given by Holec (1981, p.3), “the ability to take charge of one's own learning.” This means to take responsibility for all the decisions concerning learning. Holec's definition of learner autonomy focuses on empowering a learner to assume responsibilities traditionally taken by the teacher and also on the management and organisation of individual learning process (Benson, 2001; Little, 1991). This is similar to the “empowering technique” suggested by Kiraly (2003, pp.78-84) for a translation course, in which students are invited by the teacher to identify learning goals and contents and participate in the assessment of learning.

However, as Benson (2001) points out, Holec's definition does not make the cognitive constructs involved in learner autonomy explicit. Little (1991, p.4) argues that learner autonomy is not merely about the organisation and management of the learning process but “a capacity – for detachment, critical reflection, decision-making, and independent action.” Based on this definition, autonomy is not merely about making decisions about one's learning process, but also involves critical reflection (Candy, 1988). It is, therefore, “fundamental to autonomous learning that the learner should develop a capacity to reflect critically on the learning process, evaluate his progress, and if necessary make adjustments to his learning strategies” (Little, 1991, p.52).

The problem is that students cannot be assumed to be able to reflect critically. Helping students learn how to reflect critically should thus be an essential component of any teaching/learning method that promotes autonomy (Little, 1991, p.52).

Another essential component of learner autonomy is disciplinary knowledge. While Little (1991) believes that it should be relatively easy for an autonomous learner to transfer their autonomous capacity to other aspects of life, Candy (1988) argues that attainment of autonomy is not universal but subject-dependent. Learners' subject-matter autonomy in any particular subject area develops and accumulates when they move through stages in learning from total dependency to full autonomy. From this perspective, autonomy is not just an attitude to learning and life or an approach to learning and teaching (Boud, 1988) but is predicated upon discipline-based knowledge.

It is therefore down to a discipline to identify discipline-specific features and define levels of autonomous learning through which students' progress in autonomy can be observed and assessed. For instance, Littlewood (1996) suggests that language students' autonomy level in making decisions can be distinguished in practice: from making decisions on grammar and vocabulary to using languages independently outside the classroom. Such practice is beneficial to discipline-specific teaching, including translator training.

Reflection

Reflective ability is a desired skill of an autonomous translator, who is able to "reflect on what they do and how they do it" and "prove to themselves as well as others that they are in control of what they do" (Baker, 1992, p.4). Dewey defines reflection as "active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it, and the further conclusions to which it tends" (1910/1991, p.6). It is not casual thinking but a special form of thinking that often follows a confusing situation or is triggered by a problem, the solution of which often requires critical inquiry, review of evidence, and testing potential solutions.

While scholars like Dewey focus on the nature of reflection per se, others focus more on how reflection may bring about changes in learning (Hatton & Smith, 1995). It can be said that changes taking place serve as an indicator of reflective thinking. Such changes may occur in learning behaviour. For instance, Lee (2015), who used a reflective learning journal in a post-graduate translation practice class for a semester, reports changes in students' attitude towards translation and learning.

Reflection may also cause changes to learners' problem-solving process and perspectives of thinking. According to Schön (1983, 1987), when a professional encounters a problem that he cannot solve based on his existing "frames" of thinking, he will start to reflect on the frame itself and see if "re-framing" of the problem is required in order to solve it. The idea of frames of thinking is similar to what Habermas (1971) and Mezirow (1991) call "meaning perspectives." According to Habermas (1971), distorted meaning perspectives can be identified and challenged through self-reflection in order to emancipate oneself from instinctive, linguistic, institutional, epistemic or environmental forces that limit our options and control over our lives. Mezirow (1991, p.5) also stresses the role of reflection in assessing and changing one's own meaning schemes, which are "specific knowledge, beliefs, value judgments, or feelings" that we use to make sense of the world. If put in the context of translation, reflection is useful in helping students assess various perspectives of meaning for better decision-making. For instance, a translation decision made from a pragmatic perspective is likely to be different from a translation decision made from a functionalist perspective.

Depth of Reflection

It is generally agreed that reflection is hierarchical, but views on what defines and substantiates each level of hierarchy vary among scholars (Moon, 2004). The terms used vary and overlap to a certain degree, and sometimes are ill-defined and loosely used, leading to confusion (Hatton, et al., 1995). Nonetheless, Mezirow (1991) provides a set of well-explained definitions for levels of reflection in his transformative learning theory. He distinguishes between non-reflective and reflective action. None-reflective action is done “within pre-existing meaning schemes and perspectives” (Mezirow, 1991, p. 107). Mezirow identified three types of non-reflective action, i.e. habitual action, thoughtful action and introspection. *Habitual action* is one that we have done so frequently that we perform it without needing to give much conscious thought to it. *Thoughtful action* involves higher-orders of cognitive action, such as analysing, performing, discussing, remembering and judging, but is still based on pre-existing meaning schemes and perspectives. *Introspection* is thinking about oneself and one’s thoughts and feelings without involving validity testing of pre-existing schemes or perspectives. For instance, typing translation without giving much thought to it - such as when translating “hi” - is a habitual action. Feeling good about our translation quality is an introspection. Deciding translation quality based on prior knowledge about translation quality is a thoughtful action.

Reflective action is “making decisions or taking other action predicated upon the insights resulting from reflection” (Mezirow, 1991, p.108). It starts from an attempt to interpret an experience or solve a problem, followed by scanning, propositional construal, reflection and imaginative insight, resulting in interpretation that leads to a change or a transformation in the meaning scheme (e.g. translation beliefs) if *content reflection* (“what”) and/or *process reflection* (“how”) has occurred or, in the meaning perspective (e.g. the whole belief system), if *premise reflection* (“why”) has occurred. For instance, translation students may ask *what* problem they encounter, examine *how* they act on it, and question *why* they act as they do or even question the criteria they use to identify the translation problem in the first place. Such a categorical system of levels of reflection and their definitions may serve as a conceptual framework for assessing and analysing the depth of students’ reflection.

Assessing Reflection

With all the rich discussion on reflection as a theoretical framework, reflective thinking is still difficult to assess due to a wide range of concepts, definitions and categorisation (Rodgers, 2002; Hatton & Smith, 1995). However, efforts have been made by researchers in search of systematic, reliable and operable approaches and methods, either qualitative or quantitative.

Kember’s research team, among others, developed a readily usable questionnaire for quantitative research (Kember et al., 2000). Their questionnaire is based on a combination of literature review and initial testing and measures four constructs: *habitual action*, *understanding*, *reflection* and *critical reflection*. Except for *understanding*, the definitions of the other constructs are generally consistent with Mezirow’s (1991) transformative theory. While the *understanding* construct was initially based on Mezirow’s thoughtful action, it was subsequently narrowed down to the understanding of a concept without reflecting upon its significance in a personal or practical situation. The questionnaire was addressed to students and hence useful for investigating their own perception of levels of reflection.

Qualitative research on the assessment of levels of reflection often relies on reflective writing as a tool for data collection and then uses carefully defined coding systems for analysis, such as the one

developed by Hatton and Smith (1995). They identified four distinct types of reflective writing, namely, descriptive writing, descriptive reflection, dialogic reflection and critical reflection. *Descriptive writing* is not reflective; it simply records what occurred and makes no attempt to provide the reasons for it. *Descriptive reflection* tries to provide some reasons or justification, generally from one perspective or from the recognition of multiple perspectives. *Dialogic reflection* steps back from what occurred, has discourse with one's self, and explores possible reasons. *Critical reflection* gives reasons for decisions or what occurred taking into account the historical, social and political contexts.

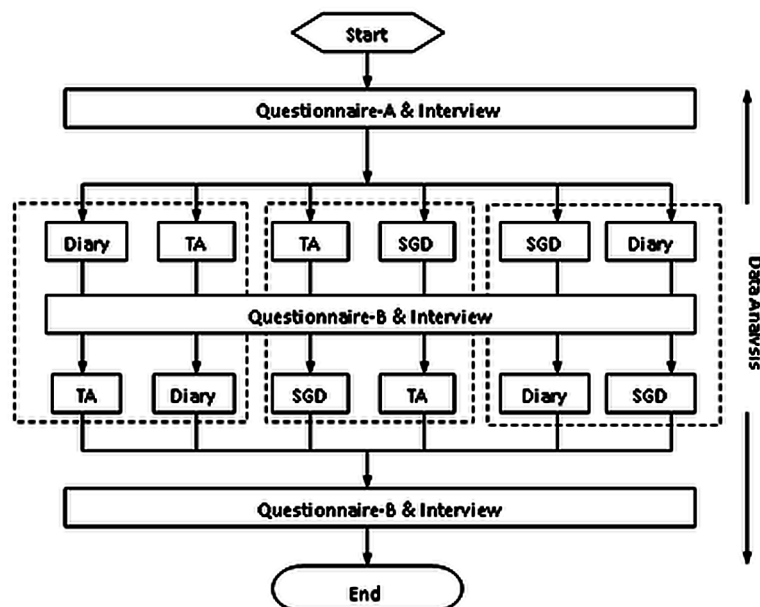
While learning diaries have been used in translator training as a qualitative method to assess students' reflective thinking, there is a scarcity of studies using quantitative approaches to this end. Translation students' reflection is usually evaluated qualitatively through reflective writing, often aiming at investigating the effectiveness of a specific reflective strategy or tool, such as Calvo's (2015) study on situated learning and Shih's (2011) and Lee's (2015) study on reflective learning journals.

METHODOLOGY

The present study adopted a qualitative and quantitative methodological design. A smaller-scale pilot study with 3 students was carried out prior to the main study to identify any deficiencies in the design. Following careful evaluation of literature, the questionnaire developed by Kember et al. (2000) and the constructs identified by Hatton and Smith were selected and tested in the pilot study and then modified accordingly for the main study in the light of students' feedback. The final design of the main study is depicted in Figure 1.

The design of the study included the following stages and tools for data collection: a learning workshop, three rounds of questionnaires, and three rounds of semi-structured interviews. The workshop

Figure 1. The design and procedure of the study



Assessing Translation Students' Reflective and Autonomous Learning

lasted around 5 months. Before its launch, an invitation explaining the purpose, procedure and ethical principles of the study was sent to the 51 Chinese students enrolled on the two-year MA in translator and interpreting (T&I) at Newcastle University. Twenty-seven students decided to participate in the workshop voluntarily, but four of them dropped out without completing the second round of questionnaire/interview. Their data were regarded as invalid and were not analysed since no data about levels of reflection were collected from them. The rest (i.e. 24 students) decided not to participate in the study.

Since the workshop ran alongside the T&I program, to avoid overburdening students, each of the participating 23 students was randomly assigned to only two of the three learning modes (i.e. Diary+TA, TA+SGD or SGD+Diary) (see section The Three Learning Modes for a full description of each learning mode). To investigate if the sequence of the two learning activities had any influence on the students' learning and thus the data gathered, students were assigned to two sub-groups in each of the three groups. For instance, in the Diary+TA group, a sub-group did Diary first and then TA while the other sub-group did TA first and then Diary. As shall be seen in the section Quantitative Results, data analysis revealed that the temporal order had no influence on the students' reflective thinking.

Before starting their first learning mode, the students filled out Questionnaire A and immediately after participated in a semi-structure interview about students' background. Each of the 23 students also completed Questionnaire B and took part in a follow-up interview after completing each learning mode. Sections "The questionnaires" and "The interviews" provide detailed information about the instruments used.

Quantitative data were then analyzed with Statistical Package for the Social Sciences (SPSS). All the qualitative data were coded and categorized according to the grounded theory approach (Charmaz, 2006).

Student Participants

The 23 participating students were Chinese (they came from either China or Taiwan) and were in their 20s or 30s. Their overall IELTS score was at least 7.0. Students did both general and specialized translation modules and translation theory modules as part of the program requirements. In theory modules, students were assessed through essays. In practical translation modules, they had weekly translation assignments and were assessed through translation as well as commentaries.

With regards to their prior experience with the three learning modes, 15 students were familiar with learning diaries and had previously kept a learning diary following teachers' instructions, including 2 students who had also kept a personal diary. Eight students had kept neither a learning diary nor a personal diary. Only eight students had heard of TA from translation theory modules but none of them were familiar with TA protocols. All of them had experience with small group discussion, which was frequently used inside and outside classroom in their MA program.

The Three Learning Modes

Each learning mode consisted of three translation tasks and a reflective learning activity (i.e. Diary, TA or STG). There was no deadline to complete the translation tasks or the learning modes except for the SGD, which required students to complete their translation before meeting for discussion in their free time.

Each student was required to translate a total of six English texts (approximately 200 words) into Chinese, three in each of the two learning modes completed by them. Fifty English texts were initially

selected from well-known newspapers or journals in the UK or the US, and then filtered down to six by four language experts to ensure comparable levels of difficulty and suitability for postgraduate level. The six source texts were randomly divided into two sets (Texts-A and Texts-B). Students translated Texts-A in the first learning mode they did and then Texts-B in the second learning mode. To investigate if levels of reflection correlated with translation quality, three translation teachers, who taught in the 2-year MA program, commented and ranked students' translations according to their quality.

In the Diary mode, each student wrote a diary entry for each of the three translation tasks they completed. In the TA mode, students were given brief training on how to think aloud while translating and they recorded themselves. Students then listened to their own TA recording and wrote a TA learning journal. In the SGD mode, students were randomly grouped in groups of three. All the SGDs were audio-recorded without the researcher's presence. The recordings were subsequently transcribed for analysis.

The following guidelines were provided for the three learning modes to serve as suggestions to students, but it was entirely up to them what to write or discuss:

- Please describe/discuss your translation process;
- Please assess/discuss your translation;
- General impression of the level of difficulty of this task;
- Others: please add anything you think is important for your learning.

The Questionnaires

Two questionnaires, A and B, were developed for the study. Questionnaire A aimed at understanding students' basic information, including age, language proficiency and previous learning experience. Questionnaire B aimed at finding out students' levels of reflection (Part I) and their views on the learning mode completed (Part II).

Part I of Questionnaire B contained 16 statements, adapted from the questionnaire developed by Kember et al. (2000, p. 395) to measure the level of students' reflective thinking in professional preparation courses. The questionnaire consisted of the four scales described in the theoretical framework (i.e. *habitual action*, *understanding*, *reflection*, and *critical reflection*). Each scale had four statements. Each statement was rated based on a 5-point rating scale, ranging from "strongly disagree" (1 point) to "strongly agree" (5 points). The total score for each scale was calculated by adding the scores of the four statements in the respective scales. If a student scored highly on a scale, it meant the student's thinking was more likely at that level of reflection.

For the purpose of the study, Kember et al.'s questionnaire was adapted. The wording of the questions was tested in the pilot study to ensure clarity and content validity. The statements are listed below. The numbers refer to the order in which each statement appeared on the questionnaire. The scores obtained were analyzed with SPSS.

Habitual Action:

1. When I am working on some translations, I can do them without thinking about what I am doing.
5. I have done translations so many times that I have started doing them without thinking about it.
9. As long as I remember what I have learned about translating, I do not have to think too much.
13. If I follow the translation skills/concepts that I know, I do not have to think too much about them.

Assessing Translation Students' Reflective and Autonomous Learning

Understanding:

2. A translation task requires me to understand translation concepts.
6. To complete a translation task, I need to understand its content.
10. I need to understand what I have learned about translating in order to perform translation tasks.
14. I have to think continually about the translation concepts/skills that I know.

Reflection:

3. I sometimes question the way I/others do translations and try to think of a better way.
7. I like to think over what I have been translating and consider alternative ways of doing it.
11. I often reflect on my translating actions to see whether I could have improved on what I did.
15. I often re-appraise my experience so I can learn from it and improve for my next performance.

Critical Reflection:

4. As a result of this learning activity I have changed the way I translate.
8. This activity has challenged some of my firmly held ideas.
12. As a result of this activity I have changed my normal way of doing translations.
16. During this activity I discovered faults in what I had previously believed to be right.

Part II of Questionnaire B consisted of the following 12 open-ended questions. In the questionnaires used for the TA and SGD learning modes, the word “diary” was replaced by “TA” and “SGD” respectively. Data collected from Questionnaires A and B served as a source of questions to be asked in the interviews.

1. What is your general impression of diary as a learning mode?
2. What are the advantages and/or disadvantages of keeping a learning diary?
3. Does keeping a learning diary help and/or hinder your reflection? In what ways?
4. How often and how long do you think it is appropriate to keep learning diaries?
5. Did knowing that you would keep a learning diary have any influence on your translating behaviours? If yes, what was the influence?
6. When did you decide what to write (e.g. while translating, before starting to write the diary, when you are writing a diary, other times, etc.)? There may be more than one answer.
7. Before you started to write your diary, did you organize your thoughts and think about what to write? If yes, how did you organize them? If no, how did you come up with a diary?
8. While you were writing your learning diary, did you come up with new ideas and/or questions? If yes, what were they? And what triggered them?
9. After you finished writing your learning diaries, have you ever thought about them? If yes, what aspects? If no, are there any reasons?
10. Did keeping a learning diary have any influence on your learning? If yes, what was the influence?
11. After completing this diary-learning mode, will you consider using diary as a learning method? If yes, why? If no, why not?
12. Please comment on your diaries as follows. How do you feel about them? Do you have any findings? If yes, what are they?

The Interviews

Three rounds of semi-structured interviews were conducted to follow up interesting findings from the questionnaires and the previous interviews. The first round of interviews aimed at finding out students' educational background, their learning motivation, difficulties and strategies, and their experience with reflective learning. The second and third rounds of interviews gathered students' perception of the learning mode they had completed, including its advantages and disadvantages and whether and how it facilitated their learning. All the interviews were audio-recorded and transcribed for analysis.

STUDENTS' PREVIOUS LEARNING EXPERIENCE

To shed some light into the interpretation of qualitative and quantitative results, students' learning experience prior to the workshop was investigated through questionnaire and interviews. It was found that all the students were intrinsically motivated and/or extrinsically motivated to learn translation. The top three motivations they reported were career considerations, interest in translation, and English enhancement.

The students' willingness to participate in the workshop, which required them to do extra translation tasks, was an additional indication that they were autonomously motivated. As one student said, "I want to try the learning methods [the researcher] mentioned (...) they might be able to help me. At least, I can do more practice [in translation]. The time will not be wasted."¹

With regards to learning, two main problems were reported by the students. First, they could not identify their own translation problems. For instance, one student said, "I think my translation is not good enough, but I don't know why and how to improve it." Second, they could not find solutions to identified translation problems. For example, a student compared herself with another female student and said, "I know my translation is not good enough and, for instance, her translation is better. But why can she translate so well? Why can't I think of those good [Chinese] sentences? Unless I could re-live her life...but that's impossible, right?" It appeared that the students felt the learning approach they were using could not solve their problems effectively, so they opted to join the workshop in search of more effective learning approaches and methods.

With regards to the use of reflection as a learning approach, all students were familiar with the meaning of "reflection", but nearly all of them perceived it as casual thinking with a focus. None of the students were familiar with the concept or definition of reflective thinking, and none of them were aware of levels of reflection.

The students' experience prior to their postgraduate study varied greatly. Some of them had experience with reflective learning activities, mostly learning diaries. The others came from teacher-centered educational environments where reflective thinking ability was seldom emphasized and students were used to getting answers directly from teachers.

QUANTITATIVE RESULTS

The quantitative analysis purported to see: (1) if the learning modes had a significant impact on students' levels of reflection, (2) if students' levels of reflection varied within a specific learning mode, and (3) if there was an observable relationship between translation quality and levels of reflection.

Assessing Translation Students' Reflective and Autonomous Learning

The descriptive statistics (Table 1) showed differences in the scores of each level of reflection across the three learning modes. However, one-way repeated measures MANOVA revealed that the three different learning modes had no statistically significant effect on students' levels of reflection since the p value was greater than .05. (Pillai's Trace = .151, $F(8, 82) = .84$, $p = .571$, $p > .05$, $\eta_p^2 = .076$).

Then, one-way repeated measures ANOVA was conducted for each of the learning modes to evaluate if there was significant difference among the four levels of reflection within a specific learning mode. In all the three learning modes, the results of the ANOVA indicated a statistically significant difference among the levels of reflection: Wilks' Lambda = .16, $F(3, 14) = 1.3$, $p < .001$, $\eta_p^2 = .845$ for the Diary mode, Wilks' Lambda = .21, $F(3, 12) = 14.7$, $p < .001$, $\eta_p^2 = .786$ for the TA mode, and Wilks' Lambda = .23, $F(3, 11) = 12.5$, $p = .001$, $\eta_p^2 = .773$ for the SGD mode. As shown in Figure 2, the order of scores in all the three learning modes was the same: *Reflection* (highest), *Understanding*, *Critical Reflection*, and *Habitual Action* (lowest). This would be indicative that doing a translation task is a cognitive activity of a higher order instead of a habitual and intuitive activity, and hence it requires students to reflect on their translation and related learning activities. However, it is noteworthy that follow-up comparisons indicated that not all the differences were significant ($p < .05$). Those pairs with significant differences are indicated by arrows in Figure 2.

As shown in Table 2, in the Diary mode, students' scores for *Reflection* were significantly higher than all the other three levels of reflection.

In the TA mode (Table 3), only scores for *Reflection* were significantly higher than *Habitual Action* and *Critical Reflection*.

Table 1. Students' level of reflection across the 3 learning modes

Descriptive Statistics				
	Learning_mode	Mean	Std. Deviation	N
HA	DIARY	10.5747	3.43407	17
	TA	9.6667	3.73529	15
	SGD	10.0000	2.21880	14
	Total	10.1037	3.17912	46
UN	DIARY	12.4171	2.43815	17
	TA	12.8000	2.51282	15
	SGD	13.7143	3.19684	14
	Total	12.9367	2.70681	46
RE	DIARY	15.4341	2.51209	17
	TA	15.4667	2.61498	15
	SGD	16.2143	3.09288	14
	Total	15.6822	2.69471	46
CR	DIARY	11.0682	3.65527	17
	TA	10.2000	2.98089	15
	SGD	13.0000	3.28165	14
	Total	11.3730	3.45592	46

Figure 2. Students' level of reflection

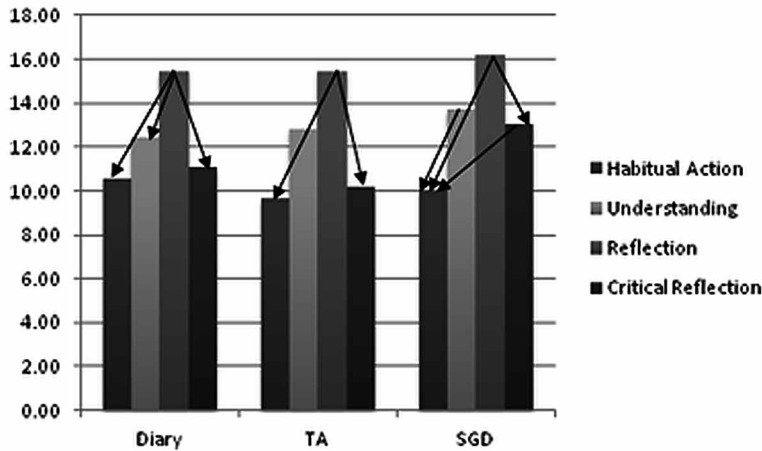


Table 2. Students' level of reflection in the Diary learning mode

Pairwise Comparisons - Diary						
Dependent Variable: Reflection Score						
(I) Reflection Level**	(J) Reflection Level	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
1	2	-1.842	1.150	.773	-5.303	1.618
	3	-4.859*	.864	.000	-7.458	-2.261
	4	-.494	1.179	1.000	-4.040	3.052
2	1	1.842	1.150	.773	-1.618	5.303
	3	-3.017*	.750	.006	-5.273	-.761
	4	1.349	.903	.927	-1.366	4.064
3	1	4.859*	.864	.000	2.261	7.458
	2	3.017*	.750	.006	.761	5.273
	4	4.366*	.699	.000	2.263	6.469
4	1	.494	1.179	1.000	-3.052	4.040
	2	-1.349	.903	.927	-4.064	1.366
	3	-4.366*	.699	.000	-6.469	-2.263

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

** . Reflection levels: 1-Habitual Action; 2-Understanding; 3-Reflection; 4-Critical Reflection

In the SGD-learning mode (Table 4), students' scores for *Reflection*, *Understanding* and *Critical Reflection* were all significantly higher than *Habitual Action*, and the *Reflection* scores were also significantly higher than *Critical Reflection*.

To understand if any relationships existed between translation quality and levels of reflection, students' scores for the four levels of reflection were compared against the ranking of their translation quality ("1"

Assessing Translation Students' Reflective and Autonomous Learning

Table 3. Students' level of reflection in the TA learning mode

Pairwise Comparisons - TA						
Dependent Variable: Reflection Score						
(I) Reflection Level**	(J) Reflection Level	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
1	2	-3.133	1.341	.209	-7.249	.982
	3	-5.800*	1.391	.006	-10.069	-1.531
	4	-.533	1.222	1.000	-4.284	3.218
2	1	3.133	1.341	.209	-.982	7.249
	3	-2.667	.887	.057	-5.390	.056
	4	2.600	1.099	.198	-.772	5.972
3	1	5.800*	1.391	.006	1.531	10.069
	2	2.667	.887	.057	-.056	5.390
	4	5.267*	.771	.000	2.900	7.634
4	1	.533	1.222	1.000	-3.218	4.284
	2	-2.600	1.099	.198	-5.972	.772
	3	-5.267*	.771	.000	-7.634	-2.900

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

** . Reflection levels: 1-Habitual Action; 2-Understanding; 3-Reflection; 4-Critical Reflection

Table 4. Students' level of reflection in the SGD learning mode

Pairwise Comparisons – SGD						
Dependent Variable: Reflection Score						
(I) Reflection Level**	(J) Reflection Level	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
1	2	-3.714*	1.102	.030	-7.137	-.292
	3	-6.214*	.944	.000	-9.149	-3.280
	4	-3.000*	.726	.007	-5.257	-.743
2	1	3.714*	1.102	.030	.292	7.137
	3	-2.500	1.020	.175	-5.670	.670
	4	.714	1.092	1.000	-2.677	4.106
3	1	6.214*	.944	.000	3.280	9.149
	2	2.500	1.020	.175	-.670	5.670
	4	3.214*	.853	.014	.565	5.864
4	1	3.000*	.726	.007	.743	5.257
	2	-.714	1.092	1.000	-4.106	2.677
	3	-3.214*	.853	.014	-5.864	-.565

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

** . Reflection levels: 1-Habitual Action; 2-Understanding; 3-Reflection; 4-Critical Reflection

representing the best quality). The two sets of texts were compared separately. As displayed in Figures 3 and 4, no consistent trends were found between the two text sets except for *Habitual Action*, where the trend line in both charts slants down to the right. The difference in scores for *Habitual Action* between the students who ranked at the top and the students who ranked at the bottom was evident. For example, in Figure 3, the student with the highest ranking (i.e. 1st) scored 13 for *Habitual Action* while the student with the lowest ranking scored only 5. They were also the students with the highest and lowest ranking respectively for Texts-B.

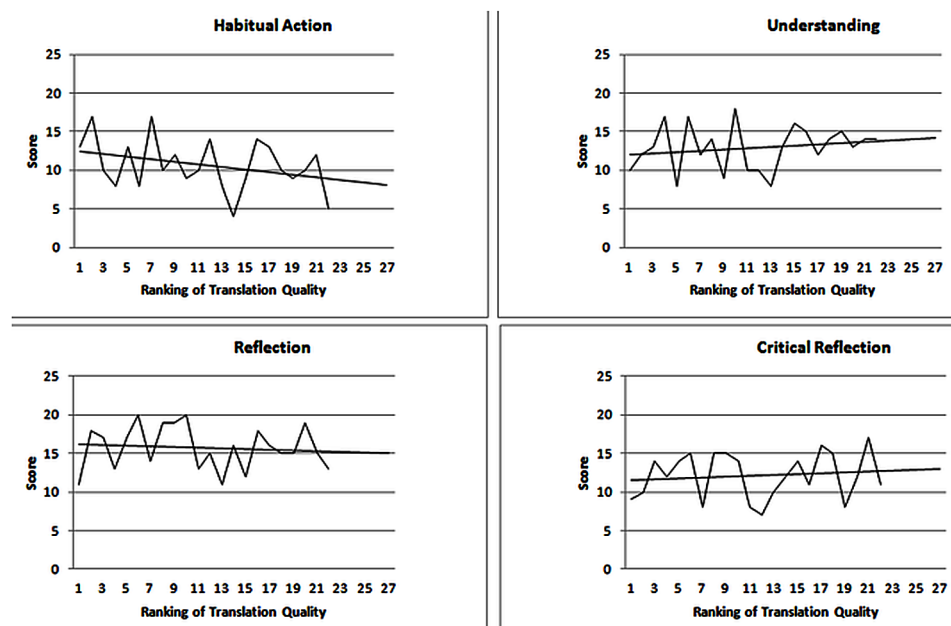
This finding that students producing translations of better quality, which is indicative of more developed translation skills, have more habitual actions could be attributed to more “routinized” and “stereotypic” decision-making processes, which occurred unconsciously without having to reflect deliberately (Prassl, 2010, p. 57).

Furthermore, the follow-up interview found that the student with the highest ranking already had 4 years of professional translation experience before enrolling in the T&I program while the student with the lowest ranking had only done some translation in her undergraduate study without any professional experience. This finding was consistent with that of Prassl’s (2010) study, where professional translators were found to have much more routinized and stereotypic decisions than students.

Qualitative Results

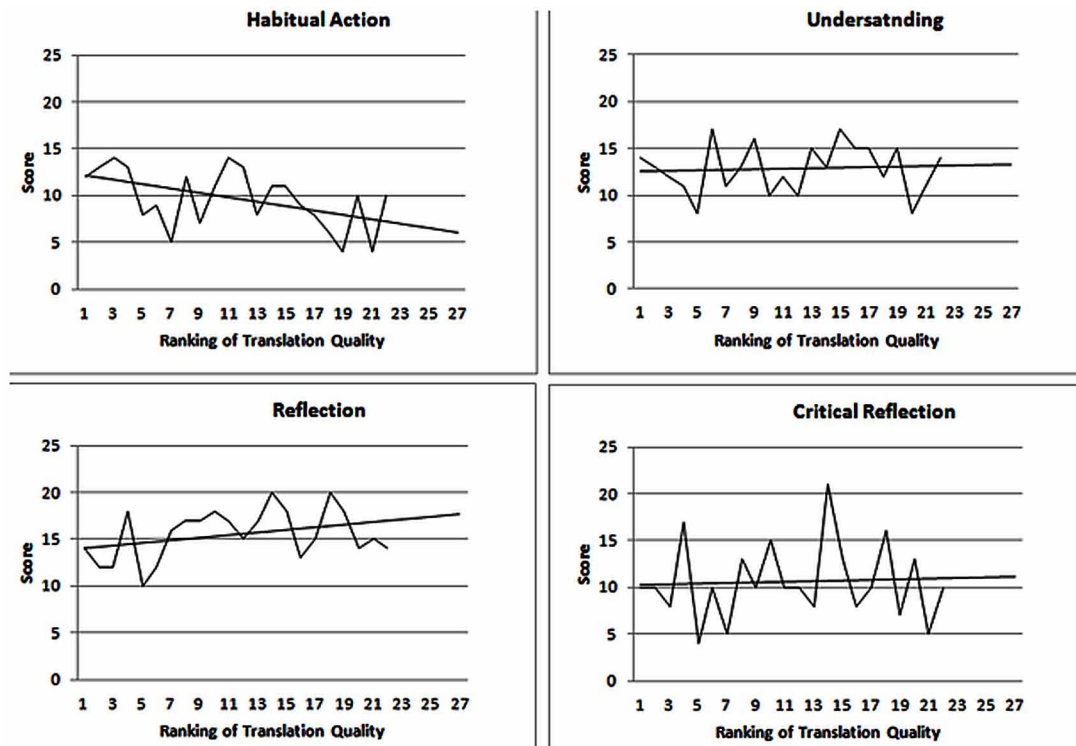
In order to identify factors influencing students’ reflective thinking and levels of reflection, the contents of the students’ diaries, TA journals and SGD transcripts were coded as *descriptions*. Each description was a complete account of an aspect, such as a procedure or problem-solving process. Nine themes emerged from the coding process: *procedure*, *problem-solving*, comments on *source text*, *translation*

Figure 3. Levels of reflection and the ranking of translation quality –Texts-A



Assessing Translation Students' Reflective and Autonomous Learning

Figure 4. Levels of reflection and the ranking of translation quality –Texts-B



quality, translation speed, overall learning, affective expression, comments on the learning modes, and others, as shown in Table 5.

In addition to the nine themes, descriptions were categorized as either *reportive* or *reflective*. A *reportive* description simply reports what occurred without attempting to give or discuss any reasons/justifications/implications for actions or behavior. An example is: “I first skimmed the whole passage, and then started to translate.” This description simply described the student’s sequence of activities.

A *reflective* description not only describes what occurred but also attempts to give reasons/justifications/implications for actions and behavior, to compare choices, to assess quality (“*lower order*” of cognitive activities), to consider multiple perspectives or questioning deeply rooted beliefs (“*higher order*”), indicating a possibility of change in students’ meaning perspectives or schemes.

The results (Table 5) showed that *problem-solving* was the most frequent theme in all of the three learning modes (39.3%, 60.5%, and 63.9% respectively) and contained much more reflective descriptions than the other themes. All descriptions coded under this theme started from something problematic, difficult or confusing, and were or were not followed by explanations, solutions or actions, depending on the level of reflection.

As for the levels of reflection, reportive description was much more frequent in the diaries and TA journals (72.9% and 77.2% respectively) whereas most SGDs were reflective (88.5%). This result was consistent with the quantitative results, where SGD scored significantly higher at the reflective level of thinking (see Figure 2). SGD also had the highest percentage of higher-order reflective descriptions

Assessing Translation Students' Reflective and Autonomous Learning

Table 5. Themes of students' reflection and percentages of levels of reflection

Learning Mode	Theme of Reflection	Reportive (%) (A)	Reflective (%)		Sub-total (%) (A)+(B)+(C)
			Lower Order (B)	Higher Order (C)	
Diary	(1)Procedure	18.4	1.7	0.0	20.1
	(2)Problem solving	24.8	14.4	0.2	39.3
	(3)Source text	10.2	4.2	0.0	14.4
	(4)Translation quality	5.8	2.3	0.2	8.3
	(5)Translation speed	9.2	0.4	0.0	9.6
	(6)Overall learning	0.0	3.1	0.6	3.6
	(7)Affective expression	2.5	0.0	0.0	2.5
	(8)Learning mode	2.1	0.0	0.0	2.1
	(9)Others	0.0	0.0	0.0	0.0
	Sub-total of (1)+(2)+..+(9)	72.9	26.1	1.0	100.0
TA	(1)Procedure	12.7	00.2	0.0	12.9
	(2)Problem solving	44.0	16.5	0.0	60.5
	(3)Source text	3.5	1.5	0.0	5.0
	(4)Translation quality	2.3	1.3	0.0	3.6
	(5)Translation speed	6.1	1.8	0.0	7.9
	(6)Overall learning	3.3	0.8	0.2	4.3
	(7)Affective expression	0.7	0.0	0.0	0.7
	(8)Learning mode	4.3	0.5	0.0	4.8
	(9)Others	0.3	0.0	0.0	0.3
	Sub-total of (1)+(2)+..+(9)	77.2	22.6	0.2	100.0
SGD	(1)Procedure	0.0	0.0	0.0	0.0
	(2)Problem solving	0.0	60.4	3.5	63.9
	(3)Source text	0.9	0.0	0.0	0.9
	(4)Translation quality	0.0	27.3	1.3	28.6
	(5)Translation speed	0.0	0.0	0.0	0.0
	(6)Overall learning	0.0	0.9	5.3	6.2
	(7)Affective expression	0.0	0.0	0.0	0.0
	(8)Learning mode	0.0	0.0	0.0	0.0
	(9)Others	0.0	0.0	0.4	0.4
	Sub-total of (1)+(2)+..+(9)	0.9	88.5	10.6	100.0

(10.6%) among the three learning modes, indicating that a collective reflective learning environment might be more likely to promote reflective thinking. Across the three learning modes, there were very few higher-order reflective descriptions, which is indicative of little change to students' existing meaning schemes and perspectives.

DISCUSSION

In addition to how students' overall reflective thinking was presented in the three learning modes, differences in reflective thinking were also found among individual students. Even at the same level of reflection, students' descriptions presented varying levels of sophistication, especially when describing their problem-solving processes. Individual students tended to maintain similar levels of sophistication throughout their learning modes and adopted a similar writing style, and the SGD groups tended to maintain similar procedures of discussion, most of which started from a problem or general comment on the translation tasks. Qualitative analyses identified several factors influencing individual students' reflective learning.

Prior Learning Experience

In addition to professional translation experience, data analysis also indicated that prior learning experience plays an important role in students' reflective thinking by influencing their frame of reference of meaning and learning approaches.

It was found that some students, when translating, looked endlessly for "accurate" meaning or teachers' answers (cf. Kiraly, 1995). They reported feeling frustrated when they could not find such meaning or answers. It is evident that these students would see the meaning of the source text and translation solutions as fixed and absolute and thus findable. The student might have inferred falsely that, if meanings were fixed, there would be no need for her to think about better solutions because the ones suggested by teachers would be the most efficient.

Such a frame of reference of meaning might have an impact on students' perceptions, and thus expectations, of how translation should be learned and taught. For instance, if students come from a transmissionist educational culture, they are more likely to believe discipline-based knowledge is transferred from teachers, so they would expect teachers to tell them what translation strategies they should use instead of thinking for themselves. For example, one student said, "The teacher[s], here [the U.K.], they do not teach you directly, they just give you concepts...to think about it later. But I have got used to the way... teachers in the mainland [China] usually taught us directly...very concrete. Sometimes I want to know the teacher's opinion. If you translate it, how do you translate that sentence? But the teachers here won't give you answers. They just give you some ideas, but I want a model to compare."

It is apparent that even if the student was no longer in a teacher-centered environment, her learning styles and attitudes were still influenced by it, which had a negative impact on her reflective learning.

Reflective Thinking Ability

Another factor emerging from data was students' reflective thinking ability, which can be evaluated from their breadth and depth of their reflection. If a student has, for instance, identified a problem but has no sufficient reflective thinking skills, he/she is unlikely to experience reflective learning.

Breadth of Reflection

Qualitative analysis showed that students who showed higher levels of reflection tended to consider the translation product or process from a wider context or multiple perspectives, such as making judgments based on readership, social and/or cultural issues rather than simply the linguistic meaning of the source text.

The following two descriptions from students' diary entries are used as examples to explain the role of theory in students' reflective activities. Both of the descriptions identified "The reddest 'red top'" as a translation problem.

Student A (Reportive): "Check new words, for example, 'redtop'? (...) looked for background information about (...) "the reddest of the red".

Student B (Reflective): "The other term "red top" was found its meaning on Wikipedia in the section which explained tabloid. So, I knew that it should refer to the nameplates of those tabloids, but I struggled whether I should translate its literal meaning "red tabloid" and added a footnote just the same as how I dealt with "spin doctors", or I should put the metaphorical meaning "tabloid". However, due to the cultural background, we did have tabloids in Taiwan, yet they were not represented in red logos. This is, the literal meaning may not share the same reflection to Taiwanese readers, so I chose to put it as "tabloid" without adding any footnote."

It is evident that the excerpts present different levels of sophistication and reflection in their analysis of how they translated the expression "The reddest 'red top'." Student A used a reportive style, simply reporting actions (i.e. looking for background information) and solutions (i.e. literal translation) without giving any explanation or justification of why she chose to translate the expression literally. Apart from the linguistic meaning of the term "red top", which may have been assimilated into the student's own meaning scheme, there was no other indication in the data that she had learned from this experience in terms of translation competence.

Student B demonstrated a much more sophisticated level of analysis and a higher level of reflection. She considered not only the linguistic meaning of "red top" but also the Taiwanese culture and readers' response. When asked why she had considered the Taiwanese culture, she answered, "we learned that in class, teachers and books about translation theories often emphasise the importance of readers' response". It is evident that, to this student, translation is not only about finding semantic equivalents but also about wider cultural contexts. Her ability to consider translation from more perspectives indicates that she has a higher level of subject-matter autonomy, which allows her to make well-informed decisions during the translation process, resulting in higher translation quality.

Student B's consideration of multi-perspectives also revealed her frame of reference of meaning and hence theoretical position towards translating and translation, as Bartrina (2005) suggests. For instance, those who only consider linguistic meaning reveal a theoretical position similar to contrastive analysis approaches, whereas those who consider both linguistic meaning and readership may have a theoretical position closer to Nord's idea of dual "loyalty" (Nord, 2000, p.185).

This implies that translation theory, particularly with regard to different frames of reference for meaning, should be an integral element of a reflective learning environment. It is reasonable to say and the data analysis of the three learning modes support the view that, if students do not have a theoretical

Assessing Translation Students' Reflective and Autonomous Learning

framework to support their reflective thinking, it is difficult for them to demonstrate deep reflection on their translating process and translation. If students do not know that the meaning of the target text is not necessarily equivalent to the meaning of the source text because meaning may be influenced by social agents or the translator's interpretation, for example, it is unlikely they would consider such influences when making translation decisions.

Furthermore, according to the concept of subject-matter autonomy (Candy, 1988; Littlewood, 1996), the better a student masters his or her discipline-based knowledge (i.e. translation theory), the more autonomous the student is in his or her discipline. Therefore, it can be inferred that students who consider translation from multiple meaning perspectives are more likely to reflect well and make their decisions wisely, leading in turn to the development of autonomy, which has an important effect on whether students will take actions to improve their translation ability and hence translation quality.

Depth of Reflection

In addition to the breadth of reflection, how deeply students can reflect also contributes to their reflective thinking ability. The results that most students reflected on translation problems instead of meaning schemes or premise of meaning schemes and that the higher-order of reflection occurred much less frequently than lower-order of reflection in the three learning modes indicated that the students might not have sufficient ability to reflect deeply. As Moon (1999, 2004) has suggested, most students do not know how to reflect critically. This was especially obvious in students' diaries and TA-journals, where most of the descriptions were reportive, not reflective. In this sense, SGD seems to be more beneficial because nearly all the SDG transcripts were reflective.

The data analysis also showed little evidence that students' meaning schemes had changed as a result of the learning modes. When students do not perceive such change, they may feel that they have not benefited from the reflective learning environment. This may partly be the reason why some of the students who did the diary and TA learning modes reported that they could not feel progress, except for having done additional translation practice.

When there is no change in their meaning schemes, students may continue to make the same or similar mistakes due to a reliance on their existing meaning schemes, which in turn may strengthen their feeling of a lack of progress. For instance, some students can be so influenced by their distorted linguistic or translation beliefs that they interpreted the source text inappropriately, made inappropriate translation decisions and even transmitted the distorted meaning schemes or perspectives to their peers in SGD (Chen, 2010).

Apart from minor changes in meaning schemes, most students appeared to have only relied on the semantic frame of reference for meaning. This narrowed down the scope and hence depth of reflection. As Livbjerg et al. (2003) and Wakabayashi (2003) have suggested, most students' thinking or reflection tends to focus on linguistic meaning.

When students do not have sufficient reflective ability, it seems less possible for them to have positive change in translation behaviour and thus translation quality. For example, Student C showed a simple, un-analytic reportive writing style in both her diary and TA-journal entries:

This was the third time I thought aloud. I still feel this is a good learning method. I am not sure the meaning of these words: invisible outlaws, invisible, The answer? (...) I don't know whether I translated them correctly. I feel background knowledge is important. I should accumulate more.

Except for recording problems, there was little evidence to show that keeping a diary or TA-journal could be of much use to her. There was also little evidence from the follow-up interview and questionnaire data to show that she had taken actions to find solutions to the problems.

In comparison, students who showed higher levels of reflection in the three learning modes tended to provide reasons or justifications for their translation decisions. They also tended to base their reasons or justifications not only on linguistic meaning, but also on readers' response and requirements of journalist articles. This indicates thus that their frame of reference of meaning comprises more than a single dimension.

Such an ability for students to reflect on the content of the translation problems (i.e. meaning) and the problem-solving process (e.g. providing justifications based on readers' response) from multiple perspectives (e.g. linguistic and dynamic equivalence) demonstrates higher levels of reflection. Such deeper and wider reflection is more likely to result in positive changes in students' existing belief systems, in the way they interpret the source text, and in translation quality. For example, one of the students reported in the interview that, after completing the Diary mode, she found herself a person "without much thought" because she was used to being guided by others not only in learning, but also in her daily and religious life:

I found I have been used to having a Truth out there to guide me...something always right out there...so I can move towards it... It will tell me what is right or wrong...just like in learning, I am used to teachers' guidance. I go to church and listen to the pastors preach.

However, after doing the Diary mode, she realized independent thinking was beneficial to her. The student stated that at least she had identified some of her translation problems and that this realization had triggered her to re-think about her learning, life and religious beliefs.

From the discussion above, it is evident that the quality of reflection is closely related to the reflective level as well as the frame of reference of meaning. The more perspectives students' frame of reference of meaning accommodates, the deeper their reflective thinking can be, and hence the better the quality of their reflection. This in turn is able to help reach the full potential of a reflective learning environment.

Awareness and Motivation

Awareness and motivation were also found to play an essential role in students' reflective thinking and hence learning. Awareness serves as a starting point of reflection. If students are not aware of translation problems, it is unlikely they will embark on a problem-solving process in the first place. Evidence was found in the three learning modes that awareness of translation problems motivated students to take the initiative in looking for solutions. It was also found, however, that some students did not take actions to solve translation problems even though they were aware of them. This implies that awareness alone is not sufficient to motivate students to take problem-solving actions autonomously. More factors are involved.

According to Littlewood (1996, 1997), the capacity for autonomy comprises both ability (i.e. knowledge and skill) and willingness (i.e. motivation and confidence). Students need to have not only motivation but also confidence, knowledge and skills to make decisions or take actions. When students perceive that all these factors are under their control, they become both more autonomous (Deci et al., 2002) and more motivated to take actions (Weiner, 1986; Williams et al., 1997).

Assessing Translation Students' Reflective and Autonomous Learning

As reported previously, students who participated in the workshop had a strong motivation to solve translation problems to develop their translation competence. Therefore, their lack of problem-solving actions despite having identified problems could be attributed, at least partially, to their low self-perceived ability or competence (e.g. disciplinary knowledge and translating skills) and low self-confidence.

Self-Perceived Competence

The data analysis revealed that some students did not take problem-solving actions despite being aware of the translation problems simply because they did not believe they had the ability to find solutions. For example, the word *literally* appeared in the three diary entries written by Student D:

When I translated the third sentence, I do not know what the author exactly means by using 'is clouded by a kind of historical amnesia', so I just looked up the meaning of "amnesia" and translated it literally.

'Deterrence optimism' means people overvalue the effect of deterrence. However, I do not know how to translate them into concise and more understandable Chinese sentences. Therefore, I just translated them literally.

I am also not sure about the meaning of 'wishful thinking', so I translated it literally as well.

When the student was asked in the survey if she had any findings, she answered, "When I cannot understand something, I will translate literally." It was evident that the reiterative use of the word "literally" in her diary made the student aware of her frequent use of literal translation as a strategy. In the post-mode interview, when asked about what this awareness meant to her, she answered, "That's it. If I knew how to improve them, I would not have translated them literally. That's the best I can do." It was apparent that despite being aware of her recurring pattern, the student did not initiate any action. According to theory of motivation, in particular the attribution theory (Weiner, 1986), when students attribute a task (i.e. finding translation solutions) as beyond their ability and perceive the attributed cause (i.e. translation competence) as internally stable and unchangeable, they tend to be less motivated to make an effort to change their status of learning. This can explain why students do not take actions to find solutions to translation problems.

Self-Perceived Confidence

The data analysis revealed that some students had actually found appropriate solutions to translation problems (which seems to indicate they were competent) but their self-confidence was too low for them to believe in their capability to find appropriate solutions. As a result, they often voiced doubts about the effectiveness of the solutions they proposed.

In comparison, students who described their translation strategies and provided reasons and justifications based on theoretical underpinning (e.g. importance of readership or genre styles) to justify their decisions to use such strategies rarely mentioned doubts about their decisions. It appears that translation theory and models provide an "explanatory power" or framework for students to make better-informed decisions (Gile, 1995, pp.13-14). This in turn endows them the confidence to justify or defend their deci-

sions. It can be inferred, therefore, that if students have acquired sufficient disciplinary knowledge and know how to apply it to defend their translation decisions, they would become more confident and perceive higher self-confidence, which, in turn, can generate more autonomous motivation to seek solutions.

Focus on Product vs. Process

Focus on translation process or translation product appears to have an effect on students' perception of self-competence and hence motivation. The data analysis showed that those who focused on translation process competencies (e.g. translating skills, speed) were more likely to perceive progress in learning; those who focused on translation product quality were less likely to perceive progress. According to motivation theory, when one perceives that one is making progress (meaning one is becoming more competent) in doing an activity, one is more likely to be autonomously motivated to do the activity (Deci et al., 2002). Therefore, it seems justifiable to infer that students who focus on the translation process are more likely to be motivated; on the contrary, when they focus more on the quality of the translation product, they are less likely to perceive increased competence and hence less likely to be motivated.

Interaction in SGD

In contrast to the Diary and TA modes, where students might or might not take actions to solve translation problems, the analysis of the SGD transcripts showed that whenever awareness of a translation problem was raised, the SGD groups always tried to find solutions even if they did not manage to settle on one. It appeared that individuals' self-perceived competence and confidence did not directly hinder the SGD groups from taking problem-solving actions, but there was evidence that student's perception of differences between their and others' competence and their affective experience had an influence on SGD members' motivation.

The data analysis showed that students are more likely to experience negative feelings such as frustration, embarrassment and shame when they perceive lower self-competence than their peers. When the influence of negative feelings is greater than the desire to improve their translation competence, students are more likely to become de-motivated. This signals that psychological cost should be taken into consideration when implementing a reflective learning activity.

For instance, one of the students admitted that she felt reluctant to engage in SGD because she felt the quality of her translations was not of the same standard as the ones produced by her SGD partners. She said, "I still forced myself to go because I wanted to learn from them." According to her SGD partners, this student distanced herself from the group after the SGD-learning mode. One of her SGD partners reported in the interview that "we used to be close friends but after experiencing the discussion learning mode in your research and, of course, the discussion in our translation class, I feel that we are not that close anymore."

The study also found that, in SGD, more competent students may experience fewer negative feelings, such as shame caused by lower perceived competence, but they may experience dissatisfaction and hence demotivation because their learning needs, such as learning good strategies from others, are not met.

A Reflection-Encouraging Learning Environment

Although no significant increase was observed in students' levels of reflection in the three learning modes, qualitative analysis showed that students perceived more reflective activities in the three learning modes than they normally would have. The reflective descriptions identified in the written data also provided evidence of reflective thinking. These all show that the three learning modes provide students with a reflection-encouraging learning environment, but they also have their own unique features, as shown in Table 6.

Awareness-Raising Effect and Problem Identification

Qualitative data revealed that the three learning modes contribute to raising students' awareness of a variety of translation problems during the translation process. This finding is consistent with prior studies. For instance, Fox (2000) found that students had heightened awareness of readers' expectations. Wakabayashi (2003) reported that her students became aware of ineffective problem-solving strategies by thinking aloud. Data analysis also showed that the Diary and TA modes help raise students' awareness of problems unknown to themselves through the process of reviewing their own diaries and TA-journals for recurring patterns or problems.

The awareness-raising effect is a valuable feature of the learning modes because it serves as a starting point of reflection. If students are not aware of their translation problems, it is unlikely they will embark on a problem-solving process in the first place.

Table 6. Comparison of the features among Diary, TA, and SGD

Features		DIARY	TA	SGD
(1)	Reflection promotion	√	√	√
(2)	Awareness-raising	√	√	√
(3)	Problem identification	√	√	√
(4)	Translation quality	√	√X	√X
(5)	Affective reaction	√	√X	√X
(6)	Use as records	√	√	X
(7)	Procedural convenience	√	X	---
(8)	Translation speed	---	√X	---
(9)	Concentration on translating	---	√X	---
(10)	Appreciation of others' translation	X	X	√
(11)	Identification of problems unknown to self	X	X	√
(12)	Exchange of views/perspectives	X	X	√
(13)	Immediate solution (or conclusion)	X	X	√X

*√: having positive effect on a feature;

X: having negative effect on a feature;

√X: having both positive and negative effects on a feature, usually depending on individuals;

---: having no apparent effect on a feature.

Design and Implementation

Although the three learning modes were found conducive to reflective learning, aspects such as design and implementation of each learning mode are key. The provision of guidelines seems to serve an important role in raising students' awareness. The interview data showed the students rarely thought about translation processes before participating in the present study and reported they would not have thought about them had it not been for the guidelines. The qualitative results also indicated that students cannot be assumed to have a good reflective ability. Like Moon (1999) suggests, interventions, such as providing a set of guidelines or teachers' facilitation, can help students to get started in reflective learning activities.

It was also found that whether students reflected on their learning or translation was greatly influenced by available time. For instance, one student said, "if I have time, I will do my best (...) to see where I went wrong or which part I should improve next time. Sometimes my time is not enough because I have to do the next translation (assignment)." This suggests sufficient time should be allowed for students to reflect when implementing a reflective learning activity.

It is also noteworthy that the SGD in the present study occurred in free-discussion groups of three, who were given no specific discussion topics except the guidelines. However, the guidelines for the SGD appeared to have had little effect on directing the SGD groups' reflections. In comparison, task-based learning settings, as described by González Davies (2004), may have the advantage that specific topics can be provided, which can direct students' reflection towards certain areas.

Nature of a Reflective Learning Activity

Students' experience with the three learning modes showed that they each have some features that might influence students' reflective learning. As shown in Table 6, diaries and TA journals were perceived by students as useful records through which they could identify recurring problems. Such actions allow them to understand their own learning, which can lead to reflective thinking and autonomous learning (Matsumoto, 1996), whereas SGD provides opportunities to appreciate others' translation, identify problems unknown to self, and enjoy benefits of cooperative learning and collective reflection, but there might be more psychological cost to consider.

Procedure-wise, the Diary and SGD modes are relatively easier than TA. Although thinking aloud appeared to be time-consuming, responses to this method varied. Some students found thinking aloud helped improve fluency or helped them concentrate on the task, leading to higher efficiency. These findings indicate that the nature of a reflective learning activity should be considered against learning goals, student needs and preferences.

Modelling Translation Students' Reflective Learning Process

Translation students' reflective and autonomous learning is a complex process. The above discussion shows that whether a reflective learning activity can help students develop reflective and autonomous learning ability depends on a number of factors. According to self-determination theory (Deci & Ryan, 2000, 2002), the pursuit of autonomy is part of human nature. It is a psychological need, just like water and air. When a learning setting enhances autonomy, better motivation may occur, which in turn may result in more satisfying learning outcomes (Vallerand et al., 2008). However, what matters is not the

Assessing Translation Students' Reflective and Autonomous Learning

setting or activity per se but how such a setting or activity can allow one to “experience feelings of competence, autonomy, and relatedness” to significant others (Vallerand et al., 2008, pp. 257).

Based on the findings of the present study, a provisional model of translation students' reflective learning process in a reflection-encouraging environment is established (Figure 5).

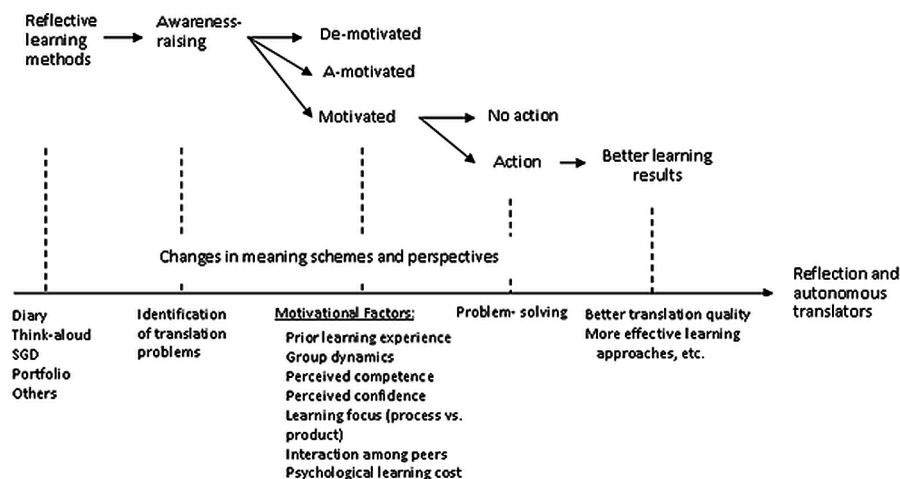
Starting from left, students first experience a reflective learning method (e.g., diary, TA and SGD), which help raise students' awareness of translation problems. Then students need to have the motivation required to take actions and then to achieve better learning outcome. Helping students move from left to right in Figure 5 is expected to help them become increasingly reflective and autonomous.

CONCLUSION

The goal of the present study was to assess students' levels of reflection in three reflective learning modes (i.e. Diary, TA, SGD). Although no statistically significant increase of students' levels of reflection in these learning modes was found, students as a group demonstrated more reflection at the Reflection level than at the Habitual Action, Understanding and Critical Reflection levels. Qualitative analyses also found students tended to adopt a reportive style in the Diary and TA modes but a predominately reflective style in SGD. The study also identified several factors that may influence students' reflective learning. Such factors include students' prior learning experience, including their cultural and educational background, learner motivation and the features, design and implementation of a reflective method. These all indicate that enhancing translation students' reflective learning requires to take into consideration students' background, learning needs and whether a reflective learning method seems to be appropriate for a given context.

This chapter has attempted to describe how students' levels of reflection and hence reflective and autonomous learning could be assessed and analyzed. The quantitative and qualitative methodological approach used proved to be very useful in identifying the trends of students' reflective thinking and hence learning. The questionnaire and the grounded-theory based qualitative analysis method were particularly

Figure 5. A provisional model of translation students' reflective learning process



useful in helping the researcher understand how and at what levels students generally reflected. Nonetheless, further research (e.g. using control groups) is necessary to have a fuller understanding of translation students' reflective learning. It is also noteworthy that the small size and generally homogeneous nature of the student participants in the present study might mean that some findings are not applicable to students from other cultural background and norms. It is hoped that further research could be done to help shed more light on translation students' reflective and autonomous learning.

REFERENCES

- Baker, M. (1992). *In other words: A coursebook on translation*. London: Routledge. doi:10.4324/9780203327579
- Bartrina, F. (2005). Theory and translation training. In M. Tennent (Ed.), *Training for the new millennium* (pp. 177–189). Amsterdam: John Benjamins. doi:10.1075/btl.60.15bar
- Benson, P. (2001). *Teaching and researching autonomy in language learning*. London: Longman.
- Boud, D. (1988). Moving towards autonomy. In D. Boud (Ed.), *Developing student autonomy in learning* (2nd ed.; pp. 17–39). London: Kogan Page.
- Calvo, E. (2015). Scaffolding translation skills through situated training approaches: Progressive and reflective methods. *The Interpreter and Translator Trainer*, 9(3), 306–322. doi:10.1080/1750399X.2015.1103107
- Candy, P. (1988). On the attainment of subject-matter autonomy. In D. Boud (Ed.), *Developing student autonomy in learning* (2nd ed.; pp. 59–76). London: Kogan Page.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. London: Sage.
- Chen, Y. (2010). Trainee translators' reflective learning in small group discussion. In V. Pellatt, K. K. Griffiths, & S. Wu (Eds.), *Teaching and testing interpreting and translating* (pp. 81–95). Bern: Peter Lang.
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268. doi:10.1207/S15327965PLI1104_01
- Deci, E. L., & Ryan, R. M. (2002). *Handbook of self-determination research*. Rochester, NY: The University of Rochester Press.
- Dewey, J. (1991). *How we think*. New York: Prometheus Books. (Original work published 1910)
- Fox, O. (2000). The use of translation diaries in a process-oriented translation teaching methodology. In C. Schäffner & B. Adab (Eds.), *Developing translation competence* (pp. 115–130). Amsterdam: John Benjamins. doi:10.1075/btl.38.12fox
- Galán-Mañas, A. (2016). Learning portfolio in translator training: The tool of choice for competence development and assessment. *The Interpreter and Translator Trainer*, 10(2), 161–182. doi:10.1080/1750399X.2015.1103108

Assessing Translation Students' Reflective and Autonomous Learning

- Gile, D. (1995). *Basic concepts and models for interpreter and translator training*. Amsterdam: John Benjamins. doi:10.1075/btl.8(1st)
- González Davies, M. (2004). *Multiple voices in the translation classroom: Activities, tasks and projects*. Amsterdam: John Benjamins. doi:10.1075/btl.54
- Habermas, J. (1971). *Knowledge and human interests*. Boston: Beacon Press.
- Hatton, N., & Smith, D. (1995). Reflection in teacher education – towards definition and implementation. *Teaching and Teacher Education*, 11(1), 33–49. doi:10.1016/0742-051X(94)00012-U
- Holec, H. (1981). *Autonomy and foreign language learning*. Oxford, UK: Pergamon Press. (first published in 1979 for the Council of Europe)
- Johnson, J. E. (2003). Learning through portfolios in the translation classroom. In B. J. Baer & G. S. Koby (Eds.), *Beyond the ivory tower: Rethinking translation pedagogy* (pp. 97–116). Amsterdam: John Benjamins. doi:10.1075/ata.xii.08joh
- Kelly, D. (2005). *A handbook for translator trainers: A guide to reflective practice*. Manchester: St Jerome Publishing.
- Kember, D., Leung, D., Jones, A., Loke, A., McKay, J., & Harrison, T. (2000). Development of a questionnaire to measure the level of reflective thinking. *Assessment & Evaluation in Higher Education*, 25(4), 370–380. doi:10.1080/713611442
- Kiraly, D. (2000). *A social constructivist approach to translator education: Empowerment from theory to practice*. Manchester: St Jerome Publishing.
- Kiraly, D. (2003). From instruction to collaborative construction: a passing fad or the promise of a paradigm shift in translator education? In B. J. Baer & G. S. Koby (Eds.), *Beyond the ivory tower: rethinking translation pedagogy* (pp. 3–27). Amsterdam: John Benjamins. doi:10.1075/ata.xii.03kir
- Kiraly, D. (2005). Situating praxis in translator education. In K. Károly & Á. Fóris (Eds.), *New trends in translation studies: In honour of Kinga Klauďy* (pp. 117–138). Budapest: Akadémiai Kiadó.
- Lee, V. (2015). A model for using the reflective learning journal in the postgraduate translation practice classroom. *Perspectives*, 23(3), 489–505. doi:10.1080/0907676X.2014.948884
- Little, D. (1991). *Learner Autonomy I: Definitions, Issues and Problems*. Dublin: Authentik.
- Littlewood, W. (1996). “Autonomy”: An anatomy and a framework. *System*, 24(4), 427–435. doi:10.1016/S0346-251X(96)00039-5
- Littlewood, W. (1997). Self-access: Why do we want it and what can it do? In P. Benson & P. Voller (Eds.), *Autonomy and independence in language learning* (pp. 79–91). London: Longman.
- Livbjerg, I., & Mees, I. M. (2003). Patterns of dictionary use in non-domain-specific translation. In F. Alves (Ed.), *Triangulating translation: Perspectives in process oriented research* (pp. 123–136). Amsterdam: John Benjamins. doi:10.1075/btl.45.11liv
- Matsumoto, K. (1996). Helping 12 learners reflect on classroom learning. *ELT Journal*, 50(2), 143–149. doi:10.1093/elt/50.2.143

- Mezirow, J. (1991). *Transformative dimensions of adult learning*. San Francisco: Jossey-Bass.
- Moon, J. A. (1999). *Reflection in learning and professional development*. London: RoutledgeFalmer.
- Moon, J. A. (2004). *A handbook of reflective and experiential learning: Theory and practice*. London: RoutledgeFalmer.
- Nord, C. (2000). Loyalty revisited: Bible translation as a case in point. *The Translator*, 7(2), 195–202.
- Orlando, M. (2011). Evaluation of Translations in the Training of Professional Translators: At the Crossroads between Theoretical, Professional and Pedagogical Practices. *The Interpreter and Translator Trainer*, 5(2), 293-308.
- Prassl, F. (2010). Translators' decision-making processes in research and knowledge integration. In S. Göpferich, F. Alves, & I. M. Mees (Eds.), *New approaches in translation process research* (pp. 57–82). Copenhagen: Samfundslitteratur.
- Rodgers, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers College Record*, 104(4), 842–866. doi:10.1111/1467-9620.00181
- Schön, D. A. (1983). *The reflective practitioner*. New York: Basic Books.
- Schön, D. A. (1987). *Educating the reflective practitioner*. San Francisco: Jossey Bass.
- Shih, C. Y. (2011). Learning from Writing Reflective Learning Journals in a Theory-based Translation Module. *The Interpreter and Translator Trainer*, 5(2), 309–324. doi:10.1080/13556509.2011.10798823
- Vallerand, R. J., Pelletier, L. G., & Koestner, R. (2008). Reflections on Self-Determination Theory. *Canadian Psychology*, 49(3), 257–262. doi:10.1037/a0012804
- Wakabayashi, J. (2003). Think-alouds as a pedagogical tool. In B. J. Baer & G. S. Koby (Eds.), *Beyond the ivory tower: rethinking translation pedagogy* (pp. 61–82). Amsterdam: John Benjamins. doi:10.1075/ata.xii.05wak
- Washbourne, K. (2014). The self-directed learner: Intentionality in translator training and education. *Perspective*, 22(3), 373–387. doi:10.1080/0907676X.2013.827225
- Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer-Verlag. doi:10.1007/978-1-4612-4948-1
- Williams, M., & Burden, R. L. (1997). *Psychology for language teachers: A social constructivist approach*. Cambridge, UK: Cambridge University Press.

ENDNOTE

- ¹ In this chapter, the original quotations by students are used despite grammar issues.

Chapter 14

The Role of Expertise in Peer Feedback Analysis: Exploring Variables and Factors in a Translation Context

Sonia Vandepitte
Ghent University, Belgium

Joleen Hanson
University of Wisconsin – Stout, USA

ABSTRACT

Since Kiraly pointed out the beneficial role of collaboration in translation training, increasingly more attention has been paid to the potential benefits of peer collaboration. While Wang and Han studied translation trainees' explicit perceptions of any benefits resulting from peer feedback, the present contribution first investigates the role of translator's implicit perceptions of reviewer expertise in the effectiveness of a peer comment in a case study. It then inquires into the number and type of peer feedback comments in relation to whether the target language that is to be reviewed is the reviewer's L1 or L2. Here, two data sets are hypothesized to yield similar results: (1) a set of native and non-native reviewer comments and (2) the comments written by translators in a direct translation situation and in an inverse translation situation. Findings were surprising, however, and professional, methodological, and theoretical research implications for translation and revision competence models are suggested.

INTRODUCTION

In the preceding decades, many didactic and pedagogical studies have pointed out the beneficial effects of collaborative work in education on the development of various transferable skills (Tudge, 1990; Berg, 1999; Paulus, 1999; Tang & Tithecott, 1999; Wells, 1999; Scardamalia, 2001; Swain, Brooks, & Tocalli-Beller, 2002; Starke-Meyerring, 2008; van Zundert, Sluijsmans, & Van Merriënboer, 2010). Recently, various translation scholars, too, have argued for the introduction of collaborative projects in

DOI: 10.4018/978-1-5225-5225-3.ch014

translation training (Kiraly, 2001; González Davies, 2004; Muñoz Martín et al., 2007; Pavlović, 2007; Kenny, 2008; Lindgren, Sullivan, Deutschmann, & Steinvall, 2009; Risku & Dickinson, 2009; Zili, 2009; Desjardins, 2011; Huertas Barros, 2011; O'Brien, 2011; Babych, Hartley, Kageura, Thomas, & Utiyama, 2012; Gambier, 2012; and OPTIMALE WP 5.4, 2013a-c). The present empirical studies, however, will focus on the process of the student's writing *act* (Toury, 2012) and, in particular, the revision phase in a translation process, which has been brought to the foreground by Mossop (2010) and Robert (2012) and Robert, Remael, and Ureel (2016). Combining both revision with peer collaboration in a translation setting has already been broached by Wang and Han (2013), who investigated students' perceptions about receiving peer feedback on their own translation, giving feedback on a peer's translation and perusing other students' work. They found that the three activities combined were especially rewarding, appreciating the alternative approach to teacher-centered education, and judging the more passive events of perceiving reviews on their own work and scanning their peers' work as more beneficial than the more active engagement of giving feedback themselves to others. A more systematic approach to collaborative translation exercise in a complete course design is described in Vandepitte (2016), who covered learning outcomes, preparatory exercises, an introduction to peer feedback, and Ghent University students' activities, who not only collaborated with each other in class, but also at home online either with each other or with students from North Dakota State University at Fargo. And Vandepitte and Gonzales (2017) have shown how the introduction of translation by peers and their feedback may lead to deeper insight into writing for different audiences and, in particular, into such linguistic matters as how to formulate definitions.

For the present studies, in which the authors inquire into whether perceptions of language expertise play an important role in student peer feedback, participants - who all signed a form of consent with regard to participating in this study - were engaged in one or two main types of collaborative project. One of the projects was an at-home project where Ghent University master translation trainees gave each other written feedback on their translation work as peers (both in a direct and an inverse translation situation). This situation was already described in Vandepitte (2016) and in Lisaité et al. (2016), who focused on the effect of student peer feedback on translation competence by applying a text-based PIE-method-based error analysis (Pre-selected Items Evaluation, Kockaert & Segers submitted) to the trainees' translations. The second project was an international Trans-Atlantic and Pacific Project (TAPP) collaboration, which provided Ghent University translation trainees with an international translation/editing setting. This project started in 1999 and more details about this project can be found in, for instance, Humbley, Maylath, Mousten, Vandepitte, and Veisblat (2005) and Vandepitte, Maylath, Mousten, Isohella, and Minacori (2016). For almost 20 years, it has connected cross-cultural virtual teams of writing, usability testing, and translation students at nearly 30 universities, 15 countries and 4 continents in a wide diversity of learning projects. Among other learning outcomes, TAPP-collaboration projects aim to prepare translation students to work in a context that adheres to industry norms BS EN-15038 (2006) and ISO 17100 (2015), involving the provision of quality assurance, including revision and review. It should be noted that the purpose of the peer feedback task was primarily didactic in a non-assessment situation, and it was provided by the students after an explicit, informative session on student feedback, teacher guidelines for which can be found in Vandepitte (2016). In order to provide a full circle feedback loop, which "usually includes a teacher who knows which skills are to be learned, and who can recognize and describe a fine performance, demonstrate a fine performance, and indicate how a poor performance can be improved" (Sadler, 1989, p. 120), students were stimulated to perform such activities in a *peer feedback* loop. From Sadler's explanation, different stages for such a peer feedback loop can be inferred. At the first stage,

The Role of Expertise in Peer Feedback Analysis

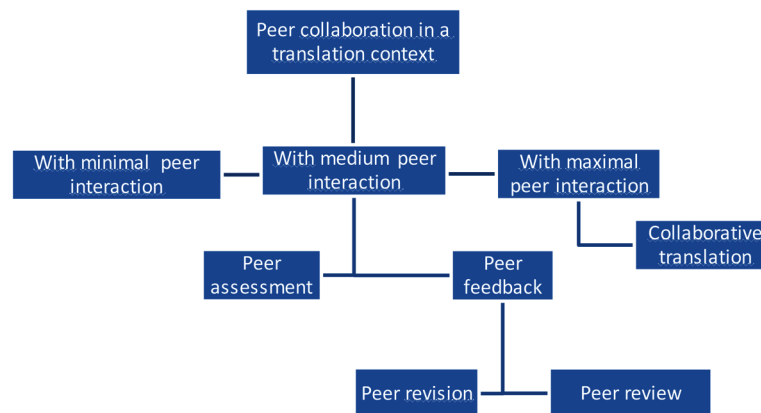
feedback givers “monitor the strengths and weaknesses” of performers so that, at a second stage “aspects associated with success or high quality can be recognized”, which can be “reinforced” at a third stage, at which “unsatisfactory aspects [can also be] modified or improved” (Sadler, 1989, p. 121). Importantly, the final stage is crucial for learning, because if “the information is simply recorded, passed to a third party who lacks either the knowledge or the power to change the outcome, or is too deeply coded (for example, as a summary grade given by the teacher) to lead to appropriate action, the control loop cannot be closed and ‘dangling data’ substitute for effective feedback” (Sadler, 1989, p. 121). It is the aim of this contribution to uncover the variables (and their factors) which stimulate students to produce quality pieces of writing/translation so that the feedback loop is completed successfully.

CONCEPTUAL BACKGROUND

In order to present the aims, methods and results of their studies, however, the authors will first present their operational definitions of the basic concepts of *feedback*, *review* and *revision* by drawing up a conceptual tree in which a node at a higher level represents a concept that includes the one below it, while the nodes connected with each other at the same level are two different hyponyms in relation to the same hypernym. With Sadler (1989), the authors see *feedback* as “a key element in formative assessment, and is usually defined in terms of information about how successfully something has been or is being done” (p. 120). In other words, it is a stage that is included in *assessment*, to wit, formative assessment, which has the aim to train and support students in an environment that provides feedback loops.

Consistent with the guiding principles of ISO 17100 (2015), the present translation training context has further made a distinction between peer comments related to the faithfulness of the translation to the source text (the issue of adequacy) - i.e. peer *revision* or, in ISO’s terms, a “bilingual examination of target language content [...] against source language content [...] for its suitability for the agreed purpose”, noting that the “term bilingual editing is sometimes used as a synonym for revision” - and those comments that are related to the idiom in the translation and how appropriate the latter is for the target audience (the issue of acceptability) - i.e. peer *review*, or in ISO’s terms, a “monolingual examination of target language content [...] for its suitability for the agreed purpose”, noting that the “term monolingual

Figure 1. A conceptual tree for peer collaboration in the translation context



editing is sometimes used as a synonym for review”. This distinction is not so easy to make for beginners of feedback exercises, though some students are very much aware of the issue, as can be inferred from the following comment made by a student as a reviewer at the end of a reviewing task, probably to explain why there were so few comments of her own:

Een aantal fouten tegen het Engels die niet echt met brontekst te maken hebben, waren al verbeterd. Dit is eigenlijk het werk voor de reviewer en niet voor de revisor. (personal comment)

(Transl: A number of errors against English that would not be related to the source text had already been corrected. This is actually the work for the reviewer, not the revisor.)

Rhetoric and composition scholars in the U.S. attach a different meaning to the term “peer review.” In U.S. composition classrooms (which focus on teaching academic writing), the term “peer review (or peer response)” refers to a specific practice that is rooted in writing process movement which revolutionized the teaching of college composition in the late 1960s and early 1970s (Elbow, 1973, 1981; Flower, 1979; Murray, 1968, 1972; Tobin & Newkirk 1994). The meaning of “peer review (peer response)” in U.S. college composition courses is different from the way “peer review” is used by translators, and it also differs from the way “peer review” is used in the scholarly publication process. Notably, in composition research and pedagogy, the terms “peer review,” “peer response,” and “peer feedback” are so similar as to be interchangeable. Little or no distinction is made between these terms. During peer review in a composition course, the goal is for student writers to provide feedback as readers, to ask questions about the meaning of the text, and to make suggestions for improving the content and organization of a classmate’s draft (Cahill, 2002; Corbett, LaFrance & Decker, 2014; Henry & Ledbetter, 2011; Mangelsdorf, 1992; McMillan, 2010; Paton, 2002; Schaffer, 1996). The purpose of peer review is to spark substantive changes to an early draft, not to correct errors in a document that is almost finished. Composition instructors strive to guide students away from focusing on the correction of surface-level errors during peer review. In contrast, peer feedback practices in a translation classroom are more likely to center on the recognition and correction of errors and to ask student to make judgments about the quality of the work (Flanagan & Heine, 2017). The “peer review” that is typical in a U.S. college composition course would not be appropriate for translation contexts in which the translator is not free to suggest adding or subtracting ideas, or radically reorganizing the sequence of ideas.

AIM

The present contribution will focus on the role of perceptions of expertise related to the target text language, which will lead to our general research question, whether perceptions of language expertise play an important role in student peer feedback. The contribution starts with a preliminary study (Study 1) that focuses on how the translator’s perception of the reviewer’s expertise influences the effectiveness of the reviewer’s comment. The authors will then draw a distinction between the translator’s perception of the reviewer’s language expertise and the reviewer’s perceptions about their own expertise. The latter may, indeed, show much variation among students: while some students only insert two or three comments, others may perform like the student whose work is shown in Figure 2.

The Role of Expertise in Peer Feedback Analysis

Figure 2. Peer review comments from a trainee with a high degree of self-perceived expertise

A come and go
Knack - 13 Nov. 2008
Theme 7/9

Rising temperatures are having an effect on mankind and the environment.

Global warming is a fact. The climate is getting more extreme, with more heat waves and floods. It is obvious that global warming has (the consequences for our country) (as well). A recent report by environmental organization WWF states that summer in European countries have been getting warmer by an average of more than 2 degrees Celsius over the last thirty years. Brussels is being somewhat spared, with an increase of 1.2°C.

Both the WWF and the Belgian Federal Council for Sustainable Development (FRDO) warn about the health problems that come with an increase in temperature. In a recent advice, the FRDO pointed out that tropical diseases such as malaria are on the rise, as is Lyme disease, which is transmitted by ticks. Pollen and dust mites, that cause allergies, are said to do better in warmer conditions. Higher temperatures cause an increased cardiovascular mortality rate anyway. The Belgian Scientific Institute for Public Health calculated that the 2003 heat wave caused an extra 1,300 senior citizens to die. On the other hand, milder winters could possibly lead to a lower death rate.

The FRDO also speculated (a little) about changes in the sea water level (An rise of 1 meter would put an additional 50,000 people of Belgian territory below sea level. One thousand years from now the sea level might, in what is called a moderate scenario, have risen by 8 meters, causing 1 tenth of Belgium to be situated below sea level. The Flemish Agriculture and Horticulture Administration recently calculated the consequences of global warming that our farmers will suffer). There was some good news. Greenhouses will need less heating, and because it is warmer some crops will be able to grow for a longer period of time. But Belgian farmers will need to take into account (the droughts that heat waves cause) heavy thunderstorms causing erosion (especially in the Flemish Ardennes and Haspengouw) and new parasites damaging their crops (such as the western corn rootworm, which is indigenous to Mexico and has already been spotted here).

It goes without saying that nature will change as well. Last spring, the organization Natuurpunt warned us, the most important developments in its periodical. (ready findings of southern insects, such as the Scarlet Dragonfly, Black-legged Bush-cricket and the Italian Striped-crow are reported on a regular basis. The Old World Bush-tit (a butterfly) also benefits from warmer, as does the Common Wall Lizard. Fifteen migratory bird species, including the Eurasian Blackcap and the Chiffchaff, are arriving at least a week earlier on (for) (spring) (spring) than for (at) (last) (year) (year). (Species like the Redstart and the Little Swift, are (at) (least) (indigenous) (birds), such as the Starling and the Blue Heron, start breeding considerably earlier on).

Plants are adapting as well. Trees are budding earlier and remain green for a longer period of time. Here and there tall trees have sprouted spontaneously. Fig trees are now becoming a normal phenomenon in our cities.

The bad news is that obviously other species are disappearing. In part because they are being replaced by the newcomers, and in part as a consequence of global warming. Our pebbles, by the way, are said to be facing hard times. A few species from the North Sea (a minuscule animal) have migrated more than a thousand kilometers north in the last twenty years. Which is important because they are a food source for quite some (fishes). (Bad prospects for the Starling (in) (the) (North) (Sea). (Common) (species) (such) (as) (the) (Atlantic) (cod) (and) (herring), and following their food (prey). Our fishermen will need to adapt their (gears). (But) (those) (that) (disappear) (will) (be) (replaced). (Recently) (a) (variety) (of) (new) (species) (of) (small) (lobsters) (and) (crabs) (are) (being) (caught) (in) (the) (North) (Sea), (and) (we) (even) (see) (the) (sea) (horse) (now)! (Some) (eighty) (new) (species) (have) (been) (registered) (so) (far).

BY DIRK DRAULANS
Copyright © 2014 Routledge Media Group. All rights reserved.

Commenting [N1]: Spaces come and go. Or: A coming and a going (of spaces)
Commenting [N2]: Translates the text
Commenting [N3]: The text is coherent and flows in a logical manner. The content has completely been translated and the register and tone have remained the same. Consequently, not a lot of revision needs to be done. I will only make a few very minor suggestions.
Commenting [N4]: You can delete this.
Commenting [N5]: In the first paragraph I must specify that it is about Belgium.
Commenting [N6]: On their website they call themselves a conservation organization.
Commenting [N7]: have got (for US: has gotten)
Commenting [N8]: Maybe better to use Fahrenheit for the US market. You write in your translation what the text is to appear in the New York Times.
Commenting [N9]: has somewhat been spared
Commenting [N10]: You have a lot of double interpreting throughout the text.
Commenting [N11]: Same as paragraph 1 (but change it to yards)
Commenting [N12]: increase is for land used for farming. Regions or areas of a country are measured in square miles.
Commenting [N13]: The consequences that our farmers will suffer (keep this clause together) because of global warming.
Commenting [N14]: This can be caused by rain (but readers know it)
Commenting [N15]: drought caused
Commenting [N16]: reported on
Commenting [N17]: put (natural) (in)
Commenting [N18]: delete this and
Commenting [N19]: At least one sea
Commenting [N20]: species are also
Commenting [N21]: not my job as
Commenting [N22]: source text says
Commenting [N23]: as is shown in 4
Commenting [N24]: See animals
Commenting [N25]: miles
Commenting [N26]: species of fish
Commenting [N27]: up north
Commenting [N28]: are being

The authors assume that such perceptions of one's own expertise will have an impact on the number and also the type of peer feedback comments they will produce. Studies 2 and 3 will, therefore, investigate peer reviews in two different situations. In Study 2, the focus will be on any distinction between the peer review feedback offered by L1 and L2 speakers of the target language to be reviewed, while Study 3, keeping participants the same, will investigate whether translation direction - into or out of the L1 - plays a role. A survey has been provided in the Appendix.

STUDY 1: EFFECTIVENESS

Since the literature has often referred to the fact that students prefer their teacher's feedback to their peers' (in a writing course, see Huahui 2010; Mangelsdorf 1992; Nelson & Carson 1998; Rollinson 2005; Yang, Badger, & Yu 2006; Zhang 1995), the *translation trainees' perceptions of reviewers' expertise* (and probably also that of revisers) may have a direct impact on the effectiveness of the comment. If translators believe that the reviewer's expertise is at least equally high as their own, they may act upon it more readily than when they estimate that the reviewer's expertise will not yield any new observations

about their own translations. Since L1 speakers are usually believed to have higher L1 expertise than L2 speakers, the authors can assume that the effectiveness of review comments from a speaker for whom the target language is the L1 (even if they do not know the source text language) will be higher than the effectiveness of review comments from a speaker for whom the target language is the L2 (and for whom the source text language is the L1).

The effectiveness of peer review comments can be operationalized as the number of comments that have been applied by the translator at the stage of finalizing the translation. In order to investigate the effectiveness of a comment, therefore, not only the comments need to be collected, but also the final products of the translator trainees and a one-on-one comparison of the student's comment with the final version needs to be carried out. Furthermore, in order to determine whether perceptions of reviewers' expertise play a role in the likelihood with which comments have been followed up, it is necessary to compare the follow-up of comments by translator trainees who have been given feedback by two reviewers whose expertise in the target language differs - preferably on the same (type of) text. Our hypothesis in this study will then be:

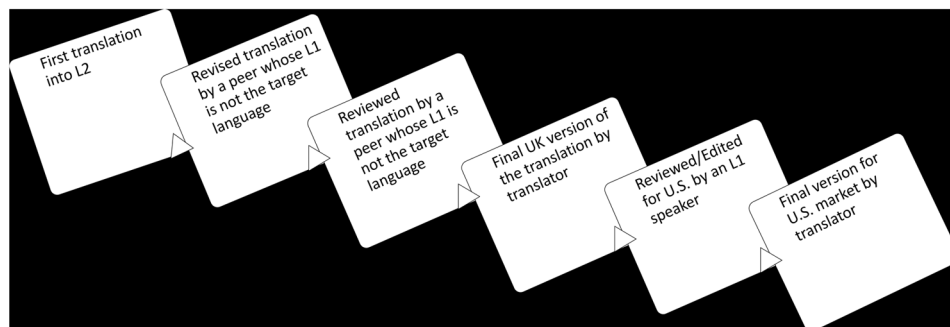
Hypothesis 1

Comments from reviewers with the target language as their L1 will be followed up to a higher degree than comments from reviewers whose L1 is not the target language

Setting

The exercise was situated within the Translation-Editing type of exercise from the Trans-Atlantic and Pacific Project carried out between Ghent University translation master's students and undergraduate students from North Dakota State University in the spring semester of 2016. It meant that one more step was added to the design that was described in Vandepitte (2016): the final translation in British English that was produced in the basic course model was sent to U.S. students for adaptation to the U.S. market. The suggestions by these U.S. students were sent to the Ghent students, who could adapt their texts further. This design is represented by Figure 3.

Figure 3. Work-flow in Study 1



Material

The source text was the Introduction to Johan Op de Beeck's *Het Verlies van België* (2015, translated by the authors as *Loss of Belgium*). It was to be translated into English for publisher Geert Cortebeek from Horizon, who presented it at the London Book Fair in the middle of April 2016. In the course of the year, it had been split up into eight parts (200-250 words each), each section being translated by four to five students. The final UK version was compiled by one student (she will be called Nadine). Nadine had taken the role of project manager for this exercise and personally selected the translated versions and finalized them into one final UK version on the basis of the comments added by her peers (ca. 30 native speakers of Dutch, L2 English as the target language). This UK version was then sent to students from NDSU, who all edited the text for the U.S. market. Only one such edited version was taken up by Nadine to finalize the translation for the U.S. market.

The data consisted of two types: 1) balloon comments from Ghent students and U.S. students; and 2) Nadine's final UK translation (in which the Ghent balloon comments from data type 1 on a translation in a target language that is the reviewers' L2 were implemented; N=27) and her final U.S. translation (in which comments by target language L1 speakers; N=31 were implemented).

Method

Nadine's two final translations were scanned to discover whether 31 U.S. reviewer comments and 27 Ghent comments were followed up. Effectiveness of comments has been operationalized as the translator's (i.e. Nadine's) positive response to a peer comment when she implemented that comment into the revised version of the translation within an agreed period of time, showing that they have implemented it. Reviewer comments were then analysed as one of the following possibilities:

- The comment has been integrated into the translation;
- The comment has been ignored;
- It is not clear whether the comment has been integrated into the translation.

However, in order for a comment to be effective in a peer feedback situation, the feedback loop must be closed in a successful way (Sadler 1989). And while a clear integration of a peer feedback comment (or a teacher's feedback comment) may lead to a better translation, it does not always guarantee an improvement of the translation. This means that an investigation into the effectiveness of comments also requires a quality judgment: students may improve the item about which the peer reviewer had raised an issue, but they may also introduce a new error, such as when they have misunderstood the comment or when their mastery of the target language is not proficient enough. Hence, an additional inquiry was carried out by the researchers into the successfulness of the final translation solution. This variable, too, was related to three factors:

- Comment follow-up is successful;
- Comment follow-up is partly successful; and
- Comment follow-up is unsuccessful.

Apart from the two analyses mentioned above - into the questions of comment implementation and its degree of success - two more questions were raised. Feedback can, indeed, even be more effective if it is also adopted in situations other than the one in which it had been made. The effectiveness of one comment may become larger if the new information that it has provided to a trainee is also applied in other passages in the text where the reviewer had not indicated any problems at all. Hence, the study also investigated whether there were any successful changes in non-commented passages. The opposite case, however, is also a possible situation - one that may even be more frequent with peer feedback than with teacher feedback: a peer reviewer's comments may still lead to a final translation that contains unsuccessful passages because remaining problems have been left undiscovered. As a result, our fourth research question was formulated as follows: Are there any remaining problems that have been left undiscovered by the peer feedback?

Results

After the analysis, it turned out that 11 out of the 31 U.S. reviewer comments (35%) were disregarded and that 9 out of 27 Ghent reviewer comments (33%) were disregarded. In other words, whether Nadine implemented peer feedback comments or not surprisingly turned out to be unrelated to whether they came from an L1 or an L2 speaker. In her case, any reviewer expertise perceptions she may have had did not have an impact on her behaviour towards the review comments: in either case, she seemed to have followed up about two thirds of the comments only.

Did her disregard for the remaining third of reviewers' comments lead to unsuccessful translations? Two passages became unsuccessful translations as a result of ignoring L1 U.S. reviewer comments, while three passages became unsuccessful translations as a result of ignoring L2 Ghent reviewer comments. These are very low figures and show that Nadine's translations were successful in, perhaps surprisingly, at least two thirds of cases where she ignored comments. The set of data is, however, too small to be able to generalize: 2 cases and 3 cases are, indeed, hardly different from one another. However, a multiplication of this ratio of one fifth (2 out of 11 unsuccessful translations) versus one third (3 out of 9 unsuccessful translations) might well be significant. So, only replication of this study with many more data points and a larger number of participants will be able to give us a more definite answer. If such a study shows similar results, generalization of the findings may well raise questions about present professional practice which only allows for target text L1 reviewers.

STUDY 2: SELF-PERCEIVED EXPERTISE - L1 VS. L2

As Study 1 has shown, it is not clear yet whether the effectiveness of comments may be related to expertise perceptions. In addition, the variable of expertise may actually play another role in peer feedback. In fact, *reviewers may have perceptions of their own expertise* and they may also have an impact on their own feedback, which may in turn have an influence on the final translation of the peer with whom the reviewer collaborates. In other words, the notion of expertise perceptions is ambiguous and needs to be disambiguated. If, indeed, the reviewers have a low degree of self-perceived expertise, they will not be very likely to offer comments about a peer's work, nor will they be likely to tackle issues that may be controversial or lead to different interpretations. If, in contrast, they have a higher degree of self-perceived expertise, they may be likely to offer many comments if they find them relevant, which may

The Role of Expertise in Peer Feedback Analysis

well enhance the final translation quality. The topic of their comments may also turn out be different, and they may also be more likely to tackle more difficult issues such as lexical nuances, textual coherence, suitability for text type and purpose. The authors hypothesize that participants with a high degree of self-perceived expertise will produce more comments and their comments will be of another kind than comments from participants with a low degree of self-perceived expertise. If the L1/L2 distinction is introduced into this study, the following hypothesis can then be formulated:

Hypothesis 2

Reviewers whose L1 is similar to the target text language will produce more comments and their comments will be of a higher order than the comments produced by reviewers whose L1 is different from the target language.

Setting

For Study 2, Ghent University translation master's students (from the same course 'English: Translation in the Foreign Language' (A703625) with the same course design, but a different group from those in Study 1) participated in a TAPP collaboration with students from the University of Wisconsin-Stout (UW-Stout) in Spring 2017. Native speakers of American English at UW-Stout provided written feedback expressing their perceptions of the correctness and readability of passages translated from Dutch to British English by Ghent translation trainees. The UW-Stout students were professional communication majors taking a course titled 'Transnational Professional Communication.' Students in this course were preparing for careers in a globalized economy where encountering language and cultural differences will be a normal part of their work with both clients and colleagues. In addition, future professional communicators can expect to play a key role in translation projects and to collaborate closely with translators (Gnecchi, Maylath, Scarpa, Mousten, & Vandepitte, 2011; Maylath, 1997; Melton, 2008; Weiss, 1999). However, despite the cultural diversity of the United States, many students who take 'Transnational Professional Communication' have not previously engaged in extended, purposeful communication with people who do not share the same L1 and cultural background. As a result, communicating with students from Ghent University about a translated text provided the U.S. students with valuable career preparation. In addition to working directly with translation trainees on a translation project, learning outcomes for the U.S. students included practicing virtual collaboration with geographically remote partners and reflecting on the impact of culture on the communication experience, including differences in work practices, politeness strategies, and English language use. The U.S. students were instructed to mark any words, phrases, or sentences that they perceived to be awkward or unnatural to American readers. They were shown an example of appropriate feedback on a sample passage that had been translated by a Ghent University student.

Material

A corpus of over 400 student feedback comments was collected from more than 50 participants, about half of whom were translation students with the source text language as their L1 and the other half were professional communication students with the target text language as their L1.

The text that the Ghent students translated was the chapter ‘Het oprukken van de vrije loonarbeid, een wereldgeschiedenis’ (this chapter authors’ translation: ‘The advance of free wage labour, a world history’) in Frank Caestecker’s *Hoe de mens de wereld vorm gaf* (2016) (this chapter authors’ translation *How mankind shaped the world*), which this Ghent University history professor provided to his students as course material in Spring 2017.

The data consisted of balloon comments in the peer reviews from both U.S. students who were instructed to use them for all the types of comments they had. Ghent students, in contrast, only applied them to issues where they might have some doubts and wanted to suggest alternatives. Even in situations where they thought it was easier to SHOW a change in the text rather than explain it in a balloon, they were requested to do so with a balloon if they were not sure that the translator would accept that suggestion. If, on the other hand, they were certain that the suggestion would be acceptable, they could use the ‘track-changes’ option and apply their *corrections* in the text itself. In other words, the commenting box comments were reserved for problem-solution only and the track-changes comments for surface changes. With this distinction students imitated local professional practice: professionals who often collaborate and revise each other’s work also respect each other’s texts, trust each other’s judgments about infelicities and may just want to easily click the ‘accept all changes’ button for lack of time to focus on the balloon comments. U.S. student comments were in English only, while the Ghent comments were either in Dutch or English or both. General comments were not included nor were the comments which were integrated in the text for matters of simplicity in the data collection: in contrast with balloons, copies of in-text revisions into an excel-file do not render clear and easily manipulatable data.

Method

All comments were analysed in terms of the issue that they dealt with. They were categorized into the lower order types of grammar (G, including morphology, word order, agreement and tense), spelling (S, including capitalization) and punctuation (P, including comments on other matters of layout) and the higher order types - which deal with meaning and macro-structural aspects of the text - of lexicon (L, including vocabulary, collocations, or other meaning-related matters) and text (T, including matters such as style, norms relevant for text purpose (e.g. avoiding repetition, short versus long forms of expression, passive versus active sentences), register, unusual phrasings, coherence and cohesion).

Results

Table 1 shows us the absolute comment frequencies for the two groups of participants with either English (EN, N:31) or Dutch (NL, N:21) as the L1 as they are distributed across the types of comment. Those figures have been normalized as a function of the number of participants for each group in Table 2:

As Table 2 demonstrates, the average number of balloon comments provided by a U.S. student was twice as high as that of a Ghent student and would confirm the hypothesis that reviewers whose L1 is similar to the target text language will produce more comments than reviewers’ whose L1 is different from the target language. The results for the second part of the hypothesis - whether reviewers whose L1 is similar to the target text language will produce *more comments of a higher order* than reviewers whose L1 is different from the target language - is also confirmed per individual student (EN L+T = 6.5 versus NL L+T = 3.4). However, this is only due to the frequency with which the U.S. reviewers have paid attention to lexical issues. The especially high frequency of comments about the lexicon per individual

The Role of Expertise in Peer Feedback Analysis

Table 1. Absolute comment frequency as a function of participant L1 language (EN or NL) and type of comment

	EN (N:31) (TL=L1)	NL (N:21) (TL=L2)	Total
G	28	12	40
S	26	9	35
P	30	8	38
L	173	50	223
T	29	22	51
Total	286	101	387

Table 2. Normalized comment frequency: average comment frequency per student as a function of participant L1 language (EN or NL) and type of comment

	EN (N:31) (TL=L1)	NL (N:21) (TL=L2)
G	0.9	0.6
S	0.8	0.4
P	1	0.4
L	5.6	2.4
T	0.9	1
Total	9.2	4.8

U.S. student (5.6 compared to 2.4 for the Ghent student) may reflect confidence in their native speaker intuition about lexical nuance (or appropriate word choice) and their higher degree of awareness of word meaning and connotation. This difference becomes even more striking when the percentages for lexical and textual comments are compared to all types of comments for each group in Table 3. This table shows that the distribution of comment types was similar within each group of reviewers (EN and NL), with approximately 30% of the feedback focused on lower order concerns (G, S, P) and approximately 70% of the comments focuses on higher order concerns (L,T) in both groups. Likewise, within both groups the most frequent comment type was the lexicon.

Table 3. Frequency of each comment type expressed as a function of the total number of comments in each participant group (TL=L1 or TL=L2)

	EN (N:31) (TL=L1)	NL (N:21) (TL=L2)
G	10%	12%
S	9%	9%
P	10%	8%
L	60%	50%
T	10%	22%
Total	100%	100%

However, the frequencies for the higher order types (L and T) are somewhat different. As expected, L1 speakers - with their higher degree of expertise - made relatively more comments about the lexicon than L2 speakers did, but, surprisingly, the L2 speakers - with a lower degree of expertise - made relatively more textual comments (on cohesion/coherence, style, etc.) than L1 speakers did. Does this point at a greater awareness of macro-structural issues among translation trainees or is it due to the instructions that were given to the U.S. students? In the PowerPoint presentation and screencast instructions that were provided to the US students, most attention went to the detection of unclear, slightly confusing, or unnatural sounding wordings. Although this may have drawn students' focus on words and phrases rather than sentences and paragraphs, none of the instructions included explicit reference to any of the word, sentence or text levels. In addition, students were given an example of how to comment on a referential infelicity. Hence, there may be another explanation why textual matters such as register, style or cohesion may not have appeared to the same extent as the other issues: since they did not understand the language of the source text, they did not know what the intended meaning of the text was and could not make any comments on content errors. However, surprisingly, the group of U.S. students *did* insert a total of fourteen comments on the content, such as the following question *I am surprised by the word "western". Wouldn't the threat be primarily from the southern, former slave-owning population?* Although this is a low number in comparison to the other types of comment issues, it shows that monolingual reviewers do want to understand the translation and match it with the knowledge framework that they already have themselves about the text topic: they made suggestions about the content based on their knowledge of U.S. history, which was the subject of the translated text. What the data do not show, however, is whether there were other content issues that U.S. students noticed, but could not or did not express, in order not to go against the source text that they could not read. If the Ghent students detected such issues, U.S. students with their higher degree of expertise would be expected to detect even more - as they did in the case of the remaining types of comments - but they did not comment on these, probably because of their lack of familiarity with the source text in Dutch. Again, this may raise questions about the professional practice of having reviews performed by monolingual reviewers only.

STUDY 3: SELF-PERCEIVED EXPERTISE - TRANSLATION DIRECTION

Study 3 is similar to Study 2 in that it also aims at comparing L1 reviewers of an English text with L2 reviewers of an English text. Setting, participants and materials, however, are different and so is the translation direction. Here, only translator reviewers participate, although they do so in two conditions: either as L2 reviewers (producing the review comments that also appeared in Study 2) or as L1 reviewers in an L2 into L1 translation task. In other words, we are comparing *comments about L1 with comments about L2 from the same participants* and will thus gain insight into the expertise perceptions that participants have about their own L1 and L2 competences. Our hypothesis is analogous to that in Study 2:

Hypothesis 3

Reviewers whose L1 is similar to the target text language - i.e. translators reviewing an L2-into-L1 translation - will produce more comments and their comments will be of a higher order than the comments produced by reviewers whose L1 is different from the target language - i.e. translators reviewing an L1-into-L2 translation.

Setting

While Study 2 covered the inverse NL-EN translation task only, Study 3 compares its Ghent University findings with peer reviews provided by a similar sample of Ghent University students in a direct EN-NL translation task. In other words, the Ghent participants do not only act as L2 reviewers, but also as reviewers for whom the target language is the L1. In the latter case, their behaviour and degree of expertise is expected to resemble that of the Wisconsin students in Study 2. While Study 2 data for the inverse translation NL-EN was collected from 21 students in Spring 2017, Study 3 also relies on comments to a direct translation exercise from EN-NL, the data of which was collected from 15 students in the preceding Autumn semester in a course with language direction EN-NL. All students in Study 3 also participated in Study 2 (some new students joined in the Spring semester).

Materials and Method

The text to be translated in this exercise was similar in length and type to that in Study 2: an extract from Chapter XXXI, The Belgian Revolution, 1830-1842, in George Edmundson's *History of Holland* (1922), which also covers a turbulent period of the reviewer's own history.

The data consisted of the same peer translator balloon reviews - which were reviews by Ghent students usually in Dutch or English about a Dutch into English translation task - as in Study 2, which was compared to the peer translator balloon reviews in an English to Dutch translation task. The latter was collected from the course 'English: Specialized Translation I (A703623)' in Autumn 2016. Again, the Ghent comments were either in Dutch or English or both. All balloon comments were analysed in terms of comment type, as in Study 2.

Results

Here, too, both absolute and normalized figures have been presented in Tables 4 and 5 respectively.

Translation trainees reviewing each other's L2-into-L1 translation (average of 3.5 per student) clearly do not produce more comments than when they review each other's L1-into-L2 translation (average of 4.8 per student). In fact, this sample shows translation trainees refraining from giving comments to each other's work in the L1. It is true that the higher order type of L2-into-L1 comments (L and T) are

Table 4. Absolute comment frequency as a function of translation direction for NL reviewers (NL-EN or NL-EN) and type of comment

	NL (N:21) (TL=L2)	NL (N:12) (TL=L1)	Total
G	12	10	22
S	9	1	10
P	8	3	11
L	50	24	74
T	22	15	37
Total	101	53	154

Table 5. Normalized comment frequency: average comment frequency per student as a function of translation direction for NL reviewers (NL-EN or NL-EN) and type of comment

	NL (N:21) (TL=L2)	NL (N:12) (TL=L1)
G	0.6	0.7
S	0.4	0.1
P	0.4	0.2
L	2.4	1.6
T	1	1
Total	4.8	3.5

more numerous than the lower-order ones (G, S, P), but this also the case for the L1-into-L2 translation direction. In addition, the inverse translation even shows a figure that is 50% higher (2.4) than that for the direct translation (1.6). It is clear that these translation students do not withhold their views on more difficult matters in the L2. Even if their expertise in the L2 is lower than in the L1, the results do not show any impact of the assumed expertise perceptions in terms of number. A qualitative investigation, however, might show more issues of uncertainty in the L1-into-L2 comments than in then L2-into-L1 comments, which is something that a future study should inquire into.

DISCUSSION

From Study 1 the authors learn that the follow-up of comments leads to better translations, and that the students showed no different response to feedback comments from TL or SL speakers. Whether there may be a slight expertise effect in favour of the TL reviewer is a matter that needs further investigation. Likewise, whether the perceived reviewer expertise may play a role in the effectiveness of particular types of comments - stylistic comments might be more readily acceptable from target language speakers than grammatical comments whose rules translation students have been trained to observe - is another matter that has not yet been studied. How could the apparent lack of impact of translators' perceptions of reviewers' expertise be explained? Perhaps the study has been guided by the wrong assumptions or there are other variables in peer reviews that play a more important role in their effectiveness. Measuring the acceptability of the final translation in comparison to the first version of the translation was useful, however, in that it may have uncovered an individual student's learning stage of development rather than just a moment in their learning path. It has also helped reveal the effectiveness of peer feedback on the development of the student's translation competence.

Studies 2 and 3 showed that perceptions of expertise may play a role in the number of comments that reviewers provided in the TAPP exercise, but that they do not seem to influence the number of comments provided by translation trainees themselves in two different translation directions. There are two possible explanations for the lower number of translation trainee comments in both studies compared to the professional communication student comments. Firstly, translation trainees may have proposed in-text changes for matters of spelling or grammar whose acceptability they were fairly certain about. Secondly, they may have shown a risk-avoidance attitude: being aware of the effort gone into a translation may make them less likely, perhaps, to produce many comments. Both reasons, however, seem to

The Role of Expertise in Peer Feedback Analysis

be contradictory: one points at confidence, the other at risk-avoidance. A study including the frequency of the in-text changes will be able to confirm whether confidence is the likelier explanation. In addition, it is quite possible that other variables such as instructions given to students may play an important role and should be included in a whole-scale investigation.

Our studies have further shown that the relation between expertise and student peer feedback in a translation context is more complex than expected. In the first place, a distinction must be made between what can be called expertise self-perceptions and expertise perceptions of the reviewer. The former refer to a person's perceptions of one's own skills to assess a peer's translation quality, whereas the latter denote a person's perceptions of a peer reviewer's skills to assess translation quality. Both types of expertise perceptions are quite different although they may interact. They may reinforce each other, as when a reviewer is quite confident in his/her own capacity of judging the acceptability of somebody else's translation and trusts the reviewers' judgements. The latter may well strengthen the translators' skill in assessing the quality of a translation product and even their own translation skills. In other cases, however, the two types of perceptions may counteract each other: a translator's (over-)confidence in his/her own translation quality assessment skills may well lead that translator to ignore a fellow student's comments, even if they turn out to be relevant and should be acted upon.

Expertise self-perceptions turn out to be in a direct relation with the number of comments a reviewer has made: the higher the expertise self-perceptions, the more comments they will make. However, the study was limited to balloon comments only, while students had also applied in-text revisions. For the translation students, those are the ones that they are most confident about, since the feedback training session taught them to do so in view of corresponding practices in the profession. Whether non-target language reviewers are assertive enough to do so, is a matter of speculation only for the time being. Sometimes, a student is, indeed, confident enough to openly express his/her disagreement with a comment raised by a peer revisor as in the following comment (referring to a peer's comment) from Study 2:

I do not agree with XXX. I believe Northern should be capitalised, as it refers to the Northern States as opposed to the Southern States. Thus, it refers to a political confederation of states.

A reliable answer, however, can only be provided by an investigation that compares those comments among both L1 and L2 reviewers.

Variables and Factors

The three studies have further provided different methodologies for peer feedback research that goes beyond Wang and Han's perception study (2013). Clearly, the methodologies need further development and refinement, in order to cross the thresholds of van Zundert et al.'s standards for good research (2010) such as experimental research with clear descriptions of methods and conditions and analytical methods with factorial designs that relate conditions to outcome variables. In particular, further research should also take into account other variables (and their factors) besides expertise that has surfaced in student feedback comments in the studies above. They can be grouped according to the following main criteria:

- **Setting:**
 - Time at which the peer feedback is provided
 - The student's history of providing peer feedback (for a particular course)

- The student's target language expertise
- The student's source language expertise
- **Materials:** Relation between target language and reviewer's L1
- **Comments:**
 - Whether in commenting boxes or as tracked changes
 - Related to revision or review
 - Comment type (G,S,P,L,T)
 - Components and their functions / relevance / clarity / evidentiality
 - Register
 - Hedging
 - Word number

For didactic purposes, for instance, special attention could be paid to the issue of relevance of a comment: does it develop positively over time with practice or does the instructor need to insert some scaffolding? While the present study focused on review comments, further inquiry into revision comments, too, would be very interesting in order to assess the adequacy/acceptability balance practices of translation trainees and whether they can be related to trainees' overall translation quality performance.

Feedback Competences

Besides the findings mentioned above, the study also had a welcome side-effect: the feedback teaching method applied in this investigation has probably not only led to the improvement of the usual transferable skills that have been witnessed in many other studies, but it may also have contributed to the development of skills in providing student peer feedback, which may be transferred in their professional life into improved translation quality assessment competence. Whether it also improved the translation trainee's desired higher quality output is a matter for future research, however.

CONCLUSION: TRANSLATION COMPETENCE AND 'REVISION' COMPETENCE

The feedback teaching method and trainees' very quick adoption of the principles of problem detection in a translation and offering of alternative phrasings raises questions about Robert et al.'s separate module for 'revision' competence (2016), a term which captures both revision and review. If, indeed, 'revision' competence is totally separate from translation competence, the authors would have expected many more problems to occur during task execution by the translation students and the writing students in our three studies. However, already in their first performance of delivering feedback comments, students - be it hesitantly for some - immediately took to the task, provided the comments to the issues they could detect and learned from the comments provided on their own and other people's translations. In that perspective, their performance would support Wang and Han's findings (2013) about translation students' perceptions about peer feedback. Perhaps, the fact that Wang and Han's students thought that providing comments themselves was less beneficial than receiving them may be explained as follows: some additional 'revision' subcompetences that do not belong to the set of translation subcompetences may demand too much effort from the students compared to the results obtained from them. However,

The Role of Expertise in Peer Feedback Analysis

the rapid speed with which the students in the above studies acquired the necessary competences necessarily points in the direction of an integrated model that captures a large number of subcompetences that both translation and 'revision' have in common, leading to a more integrative view of 'revision' competence and translation competence.

ACKNOWLEDGMENT

The authors are very grateful to Donata Lisaité for her helpful role in the collection and analysis of some data.

REFERENCES

- Babych, B., Hartley, A., Kageura, K., Thomas, M., & Utiyama, M. (2012). MNH-TT: A collaborative platform for translator training. *Translating and the Computer*, 34, 1–18.
- Berg, E. C. (1999). The effects of trained peer response on ESL students' revision types and writing quality. *Journal of Second Language Writing*, 8(3), 215–241. doi:10.1016/S1060-3743(99)80115-5
- BS EN-15038. (2006). *European quality standard. European quality standard for translation service providers*. Brussels: European Committee for Standardization.
- Caestecker, F. (2016). *Hoe de mens de wereld vorm gaf*. Ghent, Belgium: Academia Press.
- Cahill, L. (2002). Reflecting on peer review practices. In D. Roen, V. Pantoja, L. Yena, S. Miller, & E. Waggoner (Eds.), *Strategies for teaching first-year composition* (pp. 301–306). Urbana, IL: NCTE.
- Corbett, S., LaFrance, M., & Decker, T. E. (Eds.). (2014). *Peer pressure, peer power: Theory and practice in peer review and response for the writing classroom*. Southlake, TX: Fountainhead Press.
- Desjardins, R. (2011). Facebook me!: Initial insights in favour of using social networking as a tool for translator training. *Linguistica Antverpiensia. New Series—Themes in Translation Studies*, 10, 175–196.
- Edmundson, G. (2007). *History of Holland*. Charleston, SC: BiblioLife. Retrieved from <https://www.amazon.com/History-Holland-George-Edmundson/dp/1434625125>
- Elbow, P. (1973). *Writing without teachers*. Oxford, UK: Oxford University Press.
- Elbow, P. (1981). *Writing with power*. Oxford, UK: Oxford University Press.
- Flanagan, M., & Heine, C. (2017). Peer-feedback as a translation training tool in web-based communication. *Journal of Language and Communication in Business*, 54, 115–136.
- Flower, L. S. (1979). Writer-based prose: A cognitive basis for problems in writing. *College English*, 41(1), 19–37. doi:10.2307/376357
- Gambier, Y. (2012). Teaching translation / Training translators. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of Translation Studies 3* (pp. 163–171). Amsterdam: John Benjamins. doi:10.1075/hts.3.tea1

- Gnecchi, M., Maylath, B., Scarpa, F., Mousten, B., & Vandepitte, S. (2011). Field convergence: Merging roles of technical writers and technical translators. *IEEE Transactions on Professional Communication*, 54, 168–184. doi:10.1109/TPC.2011.1109/TPC.2011
- González Davies, M. (2004). *Multiple voices in the translation classroom*. Amsterdam: John Benjamins. doi:10.1075/btl.54
- Henry, J., & Ledbetter, L. (2011). Teaching intellectual teamwork in WAC courses through peer review. *Currents in Teaching and Learning*, 3(2), 4–21.
- Huahui, Z. (2010). Investigating learner's use and understanding of peer and teacher feedback on writing: A comparative study in a Chinese English writing classroom. *Assessing Writing*, 13(1), 3–17.
- Huertas Barros, E. (2011). Collaborative learning in the translation classroom: Preliminary survey results. *The Journal of Specialised Translation*, 16. Retrieved from http://www.jostrans.org/issue16/art_barros.php
- Humbley, J., Maylath, B., Mousten, B., Vandepitte, S., & Veisblat, L. (2005). Learning localization through trans-Atlantic collaboration. In G. F. Hayhoe (Ed.), *Professional communication conference proceedings, 2005. IPCC 2005. IEEE International Proceedings* (pp. 578-595). New York: IEEE. 10.1109/IPCC.2005.1494227
- ISO 17100. (2015). *Translation services: Requirements for translation services*. Geneva: ISO. Retrieved from <https://www.iso.org/standard/59149.html>
- Kenny, M. A. (2008). Discussion, cooperation, collaboration: The impact of task structure on student interaction in a Web-based translation exercise module. *The Interpreter and Translator Trainer*, 2(2), 139–16. doi:10.1080/1750399X.2008.10798771
- Kiraly, D. (2001). Towards a constructivist approach to translator education. *Quaderns: Revista de traducció*, 6, 50-53.
- Kockaert, H., & Segers, W. (2017). Evaluation of legal translations: PIE method (Preselected Items Evaluation). *JoSTrans, The Journal of Specialised Translation*, 27, 148-163. Retrieved from <http://www.jostrans.org/>
- Lindgren, E., Sullivan, K. P. H., Deutschmann, M., & Steinvall, A. (2009). Supporting learner reflection in the language translation class. *International Journal of Information Communication Technologies and Human Development*, 1(3), 26–48. doi:10.4018/jicthd.2009070102
- Lisaité, D., Maylath, B., Mousten, B., Valdez, S., Castel-Branco, M., Minacori, P., & Vandepitte, S. (2016). Negotiating meaning at a distance: Peer feedback in electronic learning translation environments in the Trans-Atlantic and Pacific Project. In M. Thelen & B. Lewandowska-Tomaszczyk (Eds.), *Translation and Meaning. Part 10. Proceedings of the Maastricht Session of the 6th International Maastricht-Lódz Duo Colloquium on "Translation and Meaning", 21-22 May 2015* (pp. 99-114). Maastricht: Zuyd University, Maastricht School of International Communication, Department of Translation and Interpreting.
- Mangelsdorf, K. (1992). Peer reviews in the ESL composition classroom: What do the students think? *ELT Journal*, 46(3), 274–284. doi:10.1093/elt/46.3.274

The Role of Expertise in Peer Feedback Analysis

- Maylath, B. (1997). Writing globally: Teaching the technical communication student to prepare documents for translation. *Journal of Business and Technical Communication*, 11(3), 339–352. doi:10.1177/1050651997011003006
- McMillan, L. (2010). Students write to students about writing. In J. Harris, J. D. Miles, & C. Paine (Eds.), *Teaching with student texts: Essays toward an informed practice* (pp. 88–95). Logan, UT: Utah State University Press. doi:10.2307/j.ctt4cgm3.13
- Melton, J. H. (2008). Lost in translation: Professional communication competencies in global training contexts. *IEEE Transactions on Professional Communication*, 52(2), 198–214. doi:10.1109/TPC.2008.2000346
- Mossop, B. (2010). *Revising and editing for translators*. Manchester: St. Jerome.
- Muñoz Martín, R., Gervilla, E., Vargas-Quesada, B., Senso-Ruiz, J., Murillo-Melero, M., Castro-Prieto, M., ... Lerma, J. (2007). A professional approach to translator training (PATT). *Meta*, 52(3), 517–528. doi:10.7202/016736ar
- Murray, D. M. (1968). *A writer teaches writing: A practical method of teaching composition*. Boston: Houghton Mifflin.
- Murray, D. M. (1972). Teach writing as a process not product. *The Leaflet*, 11-14.
- Nelson, G. L., & Carson, J. G. (1998). ESL students' perceptions of effectiveness in peer response groups. *Journal of Second Language Writing*, 7(2), 113–131. doi:10.1016/S1060-3743(98)90010-8
- O'Brien, S. (2011). Collaborative translation. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of translation studies*, 2 (pp. 17–20). Amsterdam: John Benjamins. doi:10.1075/hts.2.col1
- Op de Beeck, J. (2015). *Het Verlies van België*. Antwerpen: Horizon.
- OPTIMALE WP 5.4. (2013a). *Quality assurance*. Retrieved from <http://translator-training.eu/training/new-tools-and-technologies/quality-assurance>
- OPTIMALE WP 5.4. (2013b). *Status quo report, WP 5 – Training for professional practice, WP 5.4 – Quality Assurance*. Retrieved from http://www.ressources.univ-rennes2.fr/service-relations-internationales/optimale/attachments/article/15/WP5.4%20Synthesis%20report_EN.pdf
- OPTIMALE WP 5.4. (2013c). *Case studies of good practice in translator training - QA in the translation market*. Retrieved from http://www.ressources.univ-rennes2.fr/service-relations-internationales/optimale/attachments/article/15/WP5.4%20Case%20study_EUATC_EN.pdf
- Paton, F. (2002). Approaches to productive peer review. In D. Roen, V. Pantoja, L. Yena, S. Miller, & E. Waggoner (Eds.), *Strategies for teaching first-year composition* (pp. 301–306). Urbana, IL: NCTE.
- Paulus, T. M. (1999). The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing*, 8(3), 265–289. doi:10.1016/S1060-3743(99)80117-9
- Pavlović, N. (2007). *Directionality in collaborative translation processes* (Doctoral dissertation). Retrieved from <http://www.tesisenxarxa.net/TDX-1210107-172129/>

- Risku, H., & Dickinson, A. (2009). Translators as networkers: The role of virtual communities. *Hermes*, 42, 49–70.
- Robert, I. (2012). *La révision en traduction: les procédures de révision et leur impact sur le produit et le processus de révision*. Antwerp: University of Antwerp. Retrieved from http://www.alineremael.be/data/These_ISBN_20120425_BW.pdf
- Robert, I., Remael, A., & Ureel, J. (2016). Towards a model of translation revision competence. *The Interpreter and Translator Trainer*, 10, 1–19.
- Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT Journal*, 59(1), 23–30. doi:10.1093/elt/cci003
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144. doi:10.1007/BF00117714
- Scardamalia, M. (2001). Big change questions: Will educational institutions, within their present structures, be able to adapt sufficiently to meet the needs of the information age? *Journal of Educational Change*, 2(2), 171–176. doi:10.1023/A:1017982102060
- Schaffer, J. (1996). Peer response that works. *Journal of Teaching Writing*, 15(1), 81–89.
- Starke-Meyerring, D. (2008). *Cross-boundary knowledge making in globally networked learning environments*. Paper presented at SUNY Center for Online Collaborative Learning (COIL), Purchase, NY.
- Swain, M., Brooks, L., & Tocalli-Beller, A. (2002). Peer-peer dialogue as a means of second language learning. *Annual Review of Applied Linguistics*, 22, 171–185. doi:10.1017/S0267190502000090
- Tang, G. M., & Tithecott, J. (1999). Peer response in ESL writing. *TESL Canada Journal*, 16(2), 20–38. doi:10.18806/tesl.v16i2.716
- Tobin, L., & Newkirk, T. (Eds.). (1994). *Taking stock: The writing process movement in the '90s*. Portsmouth, NH: Heinemann.
- Toury, G. (2012). *Descriptive translation studies and beyond* (Revised edition). Amsterdam: John Benjamins. doi:10.1075/btl.100
- Tudge, J. R. H. (1990). Vygotsky, the zone of proximal development, and peer collaboration: Implications for classroom practice. In L. C. Moll (Ed.), *Vygotsky and education: Instructional implications and applications of sociohistorical psychology* (pp. 155–172). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781139173674.008
- van Zundert, M., Sluijsmans, D., & Van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, 20(4), 270–279. doi:10.1016/j.learn-instruc.2009.08.004
- Vandepitte, S. (2016). Collaboration in translation practices: A course design for peer feedback in translation training. *Konin Language Studies*, 4(4), 289–304. Retrieved from <http://ksj.pwsz.konin.edu.pl>

The Role of Expertise in Peer Feedback Analysis

Vandepitte, S., & Gonzales, L. (2017). *The language of definitions: conceptualizing technical communication with the eyes of a translation student*. Paper presented at Interdisciplinary knowledge-making: challenges for LSP research, 21st Conference on Language for Specific Purposes, Bergen, Norway.

Vandepitte, S., Maylath, B., Mousten, B., Isohella, S., & Minacori, P. (2016). Multilateral collaboration between technical communicators and translators: A case study on new technologies and processes. *The Journal of Specialised Translation*, 26, 3-17. Retrieved from http://www.jostrans.org/issue26/art_Vandepitte.pdf

Wang, K., & Han, C. (2013). Accomplishment in the multitude of counsellors: Peer feedback in translation training. *The International Journal for Translation and Interpreting*, 5(2), 62–75.

Weiss, T. (1999). The implications of translation for professional communication. In C. R. Lovitt & D. Goswami (Eds.), *Exploring the rhetoric of international professional communication: An agenda for teachers and researchers* (pp. 277–292). New York: Baywood. doi:10.2190/ETRC12

Wells, G. (1999). *Dialogic inquiry: Toward a sociocultural practice and theory of education*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511605895

Yang, M., Badger, R., & Yu, Z. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing*, 15(3), 179–200. doi:10.1016/j.jslw.2006.09.004

Zhang, S. (1995). Reexamining the affective advantage of peer feedback in the ESL writing class. *Journal of Second Language Writing*, 4(3), 209–222. doi:10.1016/1060-3743(95)90010-1

Zili, D. (2009). An empirical study on campus internet-based peer review and interactive translation. *Chinese Translators Journal*, 3, 44–49.

APPENDIX

Table 6. Study survey (EN: English, NL: Dutch)

	Study 1	Study 2	Study 3	Total
Participants	10 incl. Nadine	51	35	63
Comments	58	408	166 (NLEN 110 / ENNL 56)	466
ST Title	Verlies van België	Vrijheid	Vrijheid / History of Holland	
Source Texts (ST)	8	3	3NL / 3EN	11NL / 3EN
Target Texts (TT)	16 incl. 2 by Nadine	9	9EN / 8NL	25EN / 8NL
ST length (wd no.)	1679	647	647NL / 573EN	2326NL / 573EN

Chapter 15

Student Peer Feedback in a Translation Task: Experiences With Questionnaires and Focus Group Interviews

Carmen Heine

Aarhus University, Denmark

ABSTRACT

Social science research methods can help shed light on students' peer feedback performance. They can also help enlighten researchers on students' reception and repercussion to feedback tasks. The operationalizability of these methods for future peer activities in Translation Didactics is examined in this chapter. Multiple peer feedback data from undergraduate Business Communication students is compared with questionnaire and interview data. The data derives from peer feedback loops and provides insights into the students' perception of working with peer feedback on a web-text localization and translation commentary task performed to mirror professional practice. The analysis of the wording of student feedback, of the revisions suggested and the analysis of whether or not—and how—students implement their peer's suggestions, allows qualitative evaluation and interpretation. The methods applied are compared and their feasibility for further research into peer feedback in Translation Studies is explored.

INTRODUCTION

Peer feedback is an emerging area in Translation Didactics. It mirrors the reality of the translation profession, where qualitative measures in the form of other-revision play a crucial role. Most of the research work on peer feedback has so far been done in other research areas, such as Educational Psychology, Writing Research and Second Language Acquisition. The theoretical base of this study stems from these fields. The theoretical discussion focuses on participants' roles in peer feedback studies, scaffolding (instruction and teacher guidance (instruction used to move students progressively forward to independence in the learning process), feedback comments, student perceptions of peer feedback tasks and the process loop of feedback (a system where student comments to their peers form an action and reaction circuit;

DOI: 10.4018/978-1-5225-5225-3.ch015

the reaction being implementation or non-implementation of the feedback provided). The threefold aim is to shed light on the scaffolded feedback loop approach taken during teaching, on the feasibility and operationalizability of the research methods applied and on student perception and repercussion of the feedback task. The research questions of the study are: how do the students perform in a scaffolded peer feedback localization and academic commentary task, when they are given the possibility to implement multiple sets of feedback, and how do they perceive the task, the process, their roles as feedback providers and receivers and learning outcomes? Insight from this project may enlighten Translation Studies didactics to enhance students' learning experiences with peer feedback to prepare them for the requirements of the translation industry.

The research methods applied in this study are textual analysis of student peer feedback comments based on a coding scheme, peer feedback questionnaires and focus group interviews. The study with 19 undergraduate students in an L2 web-based communication course focuses on a peer reviewed web-text localization task from L1 into L2 and a translation commentary in L2, while the course also included other corrective written and oral peer and teacher feedback on web-copy writing and website analysis tasks¹. The peer feedback task comprised a multiple feedback loop with additional teacher feedback and was carefully scaffolded by the teachers. The scaffolding included graduated guided assistance and task instructions. Participants provided and received multiple feedback for two tasks and implemented feedback. Upon task completion, two types of recall assignments were performed. One group completed a questionnaire and received written teacher feedback. The other group participated in focus group interviews and received oral teacher feedback. Examples of feedback types provided by peers, which unfold in the form of feedback comments, e.g. suggestions, questions, or praise, form part of the data of this study. This data is examined to illustrate whether or not peer comments are implemented by the feedback recipient in his/her final translation and translation comment. The changes made are classified into revision types. An overview of this data supports the comparison to be drawn between questionnaire answers and focus group interview comments about the peer feedback task as a whole.

BACKGROUND

Editing, revision and post-editing are part and parcel of professional practices in the translation industry and quality assessment has become an established element of translator training. Industry revision practices are mirrored in the translation classroom to prepare students for their future professions and to teach them quality criteria to exert quality control and assessment practices, such as self-assessment against criteria of collaborative practices. Peer feedback as a didactic practice reflects – at least to a certain extent – expert translators' professional assessment practice. Student peer feedback is usually regarded as a productive platform for the development of evaluative skills and for learner self-regulation. Yet, reports about experiences with peer feedback in translation, analysis of how students execute peer suggestions in their texts (implementation performance), the impact peer feedback has on translation competence and reflection about these are scarce in the Translation Studies literature. Studies by Lindgren, Sullivan, Deutschmann, and Steinvall (2009), Wang and Han (2013) and Flanagan and Heine (2015), Lisaité et al. (2016) and Vandepitte (2016) are exceptions to the rule.

Advantages and disadvantages of peer feedback in general, scaffolded and non-scaffolded approaches to feedback, group composition, feedback roles, the role of L2 and the importance of the feedback loop are areas of research interest for further scrutiny. Peer feedback as evaluative judgement (Boud, 2007)

Student Peer Feedback in a Translation Task

of work against some criteria (Nicol, 2011, p. 3) is seen as a positive element of classroom practices. It is said to engage the students, contribute to learning and enhance output quality; it helps to develop explanatory skills (Cho & MacArthur, 2010) and empowers students to become independent, self-regulated learners (Macfarlane-Dick & Nicol, 2006). Students can increase their knowledge by problem-solving via feedback comments (Y. H. Cho & K. Cho, 2011). Regarding the discursiveness of feedback, the advantage of peer feedback is that peers create a discourse more understandable than the teachers' (Falchikov, 2005; Hounsell, 1997/2005; Nicol, 2011, p. 2; Topping, 1998).

Regarding pair composition, feedback quantity and variety increases with multiple peers involved (Cho & MacArthur, 2010; Topping, 1998), resulting in proportionally more complex meaning repairs at sentence and paragraph levels. Peer feedback is welcomed by students when accompanied by teacher feedback (Caulk, 1994; Tsui & Ng, 2000). Teacher involvement in the peer feedback process appears to be a preferred option, as students question peer's language competency and the credibility of the quality of feedback by their peers as opposed to that of their teacher (Nicol, Thomson, & Breslin, 2014; Tsui & Ng, 2000; Wang, 2014) and are concerned about vague comments (Mangelsdorf & Schlumberger, 1992; Min, 2005). Teacher involvement, e.g., as a last instance in a closed feedback loop, can ensure that students benefit from both their peers' suggestions of improvement and receive a final evaluation from their teacher. Teacher workload might be decreased by peer feedback tasks. Yet, introducing and incorporating the feedback loop in the classroom practice requires careful planning regarding, e.g., scaffolding, role definitions, group composition, and extra administrative workload (Nicol, 2011, p. 1). An under-researched element of the loop is the question of whether or not the teacher should not only comment on or mark the assignment upon feedback implementation, but also the feedback comments. Teachers might mark the peer comments provided by peers to encourage engagement, but this also needs to be handled sensitively (Nicol, 2011, p. 4). How this sensitivity is to be executed – and how feedback-on-feedback plays out in didactic settings – remains unanswered.

Another aspect of peer feedback is the type of feedback and the way it is given, which “can be differentially effective” (Hattie & Timperley, 2007, p. 81). It is often claimed that students over-focus on local errors (Tsui & Ng, 2000). When distinguishing between local and global errors, the higher-order concerns (HOCs) and lower-order concerns (LOCs), feedback comments which address higher-order concerns are understood to be the preferred option – from a teaching perspective. HOCs include content and focus, text organisation structure, development and voice, and stylistic problems, which change the meaning of the text. LOCs include areas such as spelling, grammar and the mechanics of writing (Krest, 1988, pp. 28-29). Such simple repairs or surface changes (Fitzgerald, 1992) are the type of feedback provided most often by inexperienced or untrained feedback providers. In studies from the fields of Educational Psychology and Writing Research, the tasks are mainly academic papers, typically longer than translation assignments. It is understandable that an abundance of LOCs may be perceived as particularly disturbing in longer texts, when researchers perceive macro-level issues as key concerns and more important for text improvement than LOCs. Earley, Northcraft, Lee, and Lituchy (1990) argue in favour of LOCs in peer feedback and Hattie, Gan and Brooks (2011, p. 304) state that they “can assist in improving task confidence and self-efficacy, which in turn provides resources for more effective and innovative information and strategy searching”. This author claims that freeing up such resources is arguably important for translation trainees that are asked to translate into a foreign language. This author argues that in translation tasks linguistic accuracy, including spelling and grammar and indeed also mechanics, play a major role regarding the target text quality. Therefore, surface comments should also be appreciated as valuable feedback.

Hattie and Timperley (2007) distinguish three types of feedback: *feed up*, which examines the learning outcomes against course content, *feed back*, which describes progress against set goals, and *feed forward*, which suggests improvement of performance for subsequent tasks. The concepts have been translated into recommendations for students (see, e.g., McMillan & Weyers, 2011). In this context, feedback is explicitly understood as positively underpinning learning. Yet, feedback is not necessarily always a reinforcer, because it can be accepted, modified or rejected (Kulhavy, 1977).

Results from studies on student perceptions of whether giving or receiving feedback is more beneficial for their learning differ. In Cho and MacArthur's (2010) study, students preferred receiving feedback; in another study, students perceived reviewing others' work as more helpful for their learning than being reviewed (Tsui & Ng, 2000). Producing reviews for peers leads to greater improvements of written assignments than receipt of reviews from peers (Y. H. Cho & K. Cho, 2011). Nicol et al. (2014) suggest that the practice of reviewing offers considerable potential in terms of mental process triggering, critical thinking, active interpretation and application of assessment criteria, reflection and learning transfer (Nicol et al., 2014, p. 112). It is argued that providing feedback encourages active comparison with students' own work. In such a comparison, internal representations of the students own work is contrasted with others.

In terms of results and outcome, implementation or deliberate and reasoned non-implementation, reflecting critical thinking, is another key factor. While teachers assume that students handle teacher or peer feedback carefully, in reality, implementation, which includes understanding and questioning and actively implementing feedback when it is deemed helpful, is often neglected. But if the feedback loop is closed – and feedback understood and implemented correctly – the desired text quality can be reached. Therefore, the feedback loop is to be understood as a series of actions, the end of which is connected to the beginning and, in translation assignments to the translation brief. Barker and Pinard (2014) understand peer feedback as a co-constructed activity between the parties involved where output (in the form of feedback) serves as impetus for action (implementation). The “loop quality” comes from the iterative procedure applied to successively approximate to the solution of a problem pinpointed by the feedback provider.

This author argues that a multiple peer feedback loop should additionally include teacher feedback on the implemented results and teacher feedback on student feedback could also be provided. Such a double loop would guarantee the often requested “follow-up” on feedback implementation and provide the possibility to live up to Nicol's suggestion that teachers should mark peer review to encourage engagement (2011, p. 4).

Positive peer feedback learning outcomes appear to depend on the number of team members and getting peer group composition right, with no consensus so far what “right” means for the individual participant and/or in varying situations. Davidson suggested assigning teams or roles (1994). A downside might be students' questionable “equal status” (Topping, 1998). Even though students may be roughly the same age or at the same class level, their performance levels may differ significantly and individual differences may affect their perception of their own status and their performance (Van Steendam, Rijlaarsdam, Van den Bergh, & Sercu, 2014). L2 performance and content knowledge differences among students may add to this difficulty in translation tasks.

Another aspect of peer feedback is the claim that it should mirror professional practice. Nicol states that feedback supports the development of disciplinary experience (2011). Feedback is, when seen by the students as joint accomplishment of a task, a collaborative activity to train students' critical skills, as well as their collaboration skills. In addition, peer feedback assignments can help students train their

Student Peer Feedback in a Translation Task

interpersonal competences (Huertas Barros, 2011). According to Price and O'Donovan (2006), gaining experience in applying formal criteria to real instances of work practice helps facilitate their internalisation. Regarding learning for the profession, authentic peer tasks calling on knowledge and skills relevant to the discipline and simulating the kinds of peer processes that occur in professional practice are deemed particularly productive (Nicol, 2011, p. 3).

The apparent consensus is that peer feedback should not “just” be implemented in the classroom without feedback instruction in the form of task instructions or criteria to accomplish set goals, as instruction is assumed to overcome at least some of the problems specified above. Such scaffolding describes learning in a continuous, supportive context. It is deemed a crucial element in teaching with feedback (Min, 2005; Nelson & Schunn, 2009; Van Steendam et al, 2014) with the unresolved debate whether scaffolding criteria should be self-administered by the students (Leijen, 2014, p. 179) or pre-defined by the teachers (Topping, 2009).

Thompson distinguishes between three types of such instruction in writing tutorials: direct instruction, which is the most directive and comes the closest to telling students what to do; cognitive scaffolding, which is also directive to some extent but is concerned with providing structured support for helping students to generate ideas and make their own revisions; and motivational scaffolding, which consists of maintaining and possibly increasing students' engagement in the writing task and in the tutorial and of helping students develop confidence as writers (Thompson, 2009, p. 419). Generally, it is suggested that feedback tasks should be carefully scaffolded of scaffolding is for students to learn how to regulate their own learning and performance (Thompson, 2009, p. 422). While experienced and efficient learners are typically sensible to task purposes, the less experienced and less efficient learners need clear guidance and instructions to direct their learning activities. By explaining in detail what the students are expected to do and how, teachers hope to trigger students' interest and students' perceptions of utility. The former refers to the intrinsic pleasure of completing the activity, the later to the advantages the students believe will arise from completing the activity (Schiefele, 1991).

Analysis methods in peer feedback research vary with the research focus. Peer feedback is often analysed by textual analysis with error types, feedback types, revision types, and implementation types elicited and coded, mainly to provide empirical evidence for their existence and frequency in particular text types and feedback settings. Many coding schemes are based on the influential Faigley and Witte (1981) scheme, e.g. in Cho and MacArthur (2010) and Flanagan and Heine (2015).

These elements link with sender characteristics, competence levels, and pair composition. Besides statistical analysis of coded feedback data, qualitative studies exploit student feedback phrasing to include criticism, praise and the impact of motivational comments on students' perception of the feedback. As specified by Stribos, Narciss and Dünnebier (2010, p. 294), prior findings on the effects of feedback content, sender characteristics and competence levels reveal, that students' perception of feedback is important. Questionnaires, interviews and observation protocols are common research tools to gather such information, focus group interviews applied to elicit ad hoc comments regarding feedback processes are frequently used. Most of the above research approaches vary significantly in scope, text production and feedback assignment type, participant demographics and methods applied. Hence, their L1 and L2 feedback assignments differ from the average translation task. Reflective commentaries – increasingly popular in translation didactics – are uncommon in other fields. But the research interests, questions, methodology and analysis means to investigate students' peer feedback, perception and repercussion are “translatable” and require scholarly attention in the translation discipline.

DESCRIPTION OF THE STUDY

Lai (2010) suggests *process*, *product*, and *perception* as dimensions for feedback analysis. In the present study², both feedback product and student perception are in the foreground. Students' perceptions of peer feedback are examined drawing on peer feedback processes as a point of departure. Students' products are drawn in, when explanations of feedback types, revision types and implementation are relevant for the study. The didactic framework of the course, the approach taken to feedback scaffolding, the social science research methods and qualitative data analysis methods applied to answer the research questions of the study are described below.

The study draws on data from students in a web-based communication course, generated in 2016. The students were L2 writers and the majority was accustomed to translation tasks from previous courses. In addition to website analysis and web-copy writing, the course contains translation and translation reflection elements, which mirror professional translation and review practices. They also comprise students' academic reflections regarding the translation problem space (TPS), which, according to Shei, is a hypothetical mental state, which translators experience, in one form or another, when embarking on a translation task (Shei, 2005, pp. 313-314).

The assignment, a web-text localization task with accompanying detailed translation brief and task description, included a self-reflective academic review in the form of a translation commentary (Shei, 2005). The web-text localization task required the students to translate an introduction website text of the Danish sweets company "Spangsberg" into English. Non-Danish speaking students used Google Translate; Danish-speaking students translated it themselves. The assignment was intended as a full circle peer feedback loop (Sadler, 1989) in student quads, where multiple peer feedback was provided to each student. It also included a teacher feedback loop, in which the teachers provided feedback, based on the final result of the multiple peer feedback assignment, and reflected on the feedback provided by the students.

Regarding group composition, the study opted for diverse membership and the students formed groups to include one international student each. Every participant shifted roles from reviewer to feedback receiver. The teachers were aware that participant selected teams are not necessarily diverse, but given that students were from 4th and 6th semesters and from different language combinations and given that the course included international students, group formation turned out to be fairly natural. The students provided peer feedback to all members of their group, using track changes and the commenting function in Word. 19 students participated in the study in one way or other. 15 students handed in the assignment with feedback comments included. 10 students received feedback from 3 peers. In one assignment, 2 students received feedback from 2 peers and in 5 assignments a pair provided feedback for each other. 1 of the 19 participants did not hand in the assignment but participated in the interview, 4 students submitted their final documents without feedback comments. The researcher knows from their comments in questionnaire and interviews that these students have provided and received feedback.

The translation brief specified that students were expected to produce a high-quality, professional text. They were instructed to implement the feedback from their peers where they deemed it appropriate into their localized web-texts and translation commentaries and to resubmit their first localized version, all feedback comments from their group members and their final texts with feedback implemented. In addition, they submitted their translation commentaries.

The reflective commentary was to include selected elements of the translation which caused problems or which were interesting from a process perspective. The students were provided with cognitive

Student Peer Feedback in a Translation Task

scaffolding via an information sheet, which suggested a possible text structure for the translation commentary and for a comment on feedback implementation/non-implementation. The peer feedback task was also carefully scaffolded in the didactic setup of the course, with previous and subsequent oral and written feedback tasks and sessions, where feedback was provided by peers or teachers. The translation brief, the task description and the information sheet formed the explicit instructional scaffolding of the web-text localization task, to aid the students in their task interpretation and to ensure their meaningful and thoughtful engagement with the task. Implicit scaffolding was available from other written and oral feedback tasks on text products during the 14 weeks long semester and from the course descriptor. From these, the students were aware that feedback comments were expected to include linguistic, academic writing and domain content. They had also participated in developing and applying assessment and evaluation criteria derived from the course descriptor in previous assignments. Feedback as means of learning had been discussed in class; examples of concepts like e.g. feed up/feed back/feed forward and examples of feedback types, their wording and their function and malfunction (Macfarlane-Dick & Nicol, 2006) had been provided together with readings about these concepts. For this study, the teachers took their responsibility to help the students to develop the skills necessary in helpful feedback literally (Blair, Curtis, Goodwin, & Shields, 2013). While scaffolding should be specifically tailored to tasks and not be transferred across courses or task types, this course is an exception, as it draws on scaffolding approaches and task instructions from previous years. The instructions provided did not include explicit reference to the fact that the task was designed to mirror professional practice.

Textual analysis and social science research methods were applied to gather information about the students' perceptions of the task. Upon task completion, the student group was split and two methods were applied to prompt the students' reactions inquiring into feedback provision, reception and implementation experiences. The study gathered information about students' experiences, attitudes and concerns regarding the feedback task. Seven out of nineteen students completed a questionnaire based on open and closed questions directly posed upon completion of the task in an online session. The other twelve participated in focus groups on the day after submission, using the same open and closed questions and additional questions that arose from the focus group situation. Method groups were split to test the methods against each other and to see, which of the methods would allow speedy results to be reported back to the students as soon as possible after the reports. The focus group participants were assigned to two groups to ensure that participants could speak freely about their experiences and the quality of the feedback they received, as none of their feedback providers took part in their respective focus group interviews. As motivational scaffolding and in order to maintain students' engagement in the task and to ensure a full-circle feedback loop for the questionnaire group and the focus groups, the former received written teacher feedback upon completion of the questionnaire form, the latter received oral teacher feedback immediately after the interview.

METHOD DESCRIPTION

For the textual analysis of the feedback types, revision types and implementation types of this study, a coding scheme based on Cho and MacArthur (2010) and Flanagan and Heine (2015) was used. The feedback types were coded into five categories: motivational feedback, suggestions (directives), suggestions (expressives), questions and statements. Motivational feedback comprises praise by the peers in the form of comments like *"Nice!"* or *"Good solution!"* for parts of the task or for the assignment overall,

e.g. “*A very nice translation all-in-all*”. Directive suggestions include explicit suggestions (solutions) by the peer. Expressive suggestions are comments on specific aspects of writing, often providing both problem and solution. Questions are feedback comments that use a question format to highlight aspects of writing. Statements are non-directive, neutral feedback comments. The analysis of the feedback types shows how the students phrase their feedback.

The revision types comprise three types of changes: surface changes, micro-level meaning changes and macro-level meaning changes. The surface changes fall into two categories: formal changes, e.g., spelling, punctuation and tense shift suggestions, and meaning-preserving changes, e.g., additions, deletions and substitutions. The micro-level meaning changes fall into two categories: complex repairs, e.g., fixing issues by deletion or changing points in a sentence, and extended content changes, e.g., elaboration on/or justification of an existing point or example. The macro-level meaning changes also fall into two categories: new content, e.g., new points, including new paragraphs, and organisation, e.g. changes to, addition of and deletion of text, change or deletion of connectives/transitional phrases. The results of the revision type analysis can shed light on whether the feedback comments fall into the LOCs or HOCs categories.

The implementation types fall into implemented and non-implemented feedback. Implementations range from inaccurate sources with accurate feedback and accurate final text to accurate sources with inaccurate feedback and inaccurate final text. Results from the implementation type analysis can be used to interpret qualitatively whether the accuracy of feedback plays a role in its implementation. In combination, the analysis of the three types displays the width, breadth and depth of scaffolded multiple peer feedback and lends itself to data triangulation.

In order to understand learner’s general perceptions of peer feedback, questionnaires, journal logs and interviews are the classical instruments used (Chang, 2016, p. 86). Questionnaire study and the focus group interview were chosen as follow-up on the peer feedback assignment, in order to solicit information from the students and to obtain insights on areas of interest, which were defined as “providing and receiving feedback”, “feedback roles”, “feedback perception”, “perceived learning outcomes”, “sensitivity of responsibility” and as a “gratification measure”.

The questionnaire was answered during a two-hour online session in the environment of their choice. It comprised a total of twenty questions to be answered on a 5-point Likert scale. Twelve questions were closed questions, eight of the questions included additional open prompts that focused on particular aspects of the question; room was left for general comments at the end of the questionnaire. The focus group filled in the same questions as their questionnaire counterparts but were prompted by the teachers to elaborate on the open questions during the interview. For the focus group interview a permissive, non-threatening environment was created. The two methods were chosen to ensure that the teachers were able to explore student responses on the scaffolded peer feedback approach, based on different sources instead of textual analysis of feedback comments alone. The aim was to discover what was done right or wrong in teaching with peer feedback.

The qualitative and quantitative data from questionnaire questions were used as a method to allow conclusive and speedy access to the information provided. The advantage in the use of questionnaires lies in the low investment in evaluation time. A quick turnaround was relevant, because immediate teacher feedback on the feedback assignment was to be given to the students in class following the feedback assignment. Yet, instead of choosing closed questions only, open-ended questions were added to personalise, to make the tool more flexible and to allow for quantitative and qualitative analysis for comparison with the focus group interview comments.

Student Peer Feedback in a Translation Task

Focus group interviews are a qualitative research tool. Contrary to questionnaires, they are exploratory rather than conclusive. The interactive setting allowed the students to express themselves freely. The interview questions were based on the questionnaire questions, but the students were asked to elaborate on the open-ended questions voiced by the teachers. Focus group interviews allow flexibility to change the questioning during the interview and allow interesting aspects from the conversation to be discussed. The students can bounce ideas off one another and engage in an opinion-based conversation about their multiple peer feedback task and their diverse opinions and perspectives critical to the topic. Focus group interviews can provide much information in a relatively short time. This was relevant, again because an account of the interview was to be given to the students in class following the feedback assignment. Unlike the questionnaire, concerns, frustrations and specific incidents of the feedback task can be covered during the focus group interviews. In addition, Nicol et al. (2014) highlight that their focus group study did not only provide quantitative explanations but produced crucial reviews of the process of providing and receiving feedback. Focus group interviews and questionnaires can shed light on students' perceptions that might in turn help understand learning mechanisms at play.

ANALYSIS AND DISCUSSION

Textual Analysis

The textual analysis consists of 216 feedback comments, 149 comments regarding the web-localization task, and 67 comments regarding the translation commentary. The lowest total number of feedback comments given by feedback providers per recipient was 5 and the highest 32 comments. Table 1 provides an overview of the feedback type distribution.

Previous studies (Cho & MacArthur, 2010; Strijbos, Narciss & Dünnebier, 2010) state that motivational comments are generally well received with students. In this study, 22.69 per cent of comments are motivational comments (mainly in the form of praise). The relatively high number of motivational comments is in alignment with the questionnaire data regarding student perception of the feedback task. Directives (26.85 per cent), expressives (28.71 per cent) and motivational comments were high as expected, with a matching high number of formal (28.59 per cent) and meaning-preserving (51.94 per cent) surface changes from the revision types (see Table 4).

Suggestions and suggestions combined with other feedback types also invoked macro-level and micro-level meaning changes of the revision types (see Table 4). The students are not only aware of the distribution of feedback types, they can also pinpoint for which task which type of feedback would have been relevant, as the following quote from the open ended comments section of the questionnaire indicates:

Table 1. Feedback comments according to feedback type

Feedback Type	Percentage
Motivational	22.69
Suggestions (Directive)	26.85
Suggestions (Expressive)	28.71
Questions	4.62
Statements	17.13

Table 2. Results non-implemented feedback

Non-Implementation	Percentage
Non-implemented - without interpretable explanation	81.21
Non-implemented with an interpretable reason	18.79

Table 3. Results implemented and/or improved feedback

Implementation Categories	Percentage	
Implemented		
Yes (implemented) in a note	5.16	
Yes (implemented) Source inaccurate, feedback accurate, final text inaccurate	6.45	
Yes (implemented) source inaccurate, feedback inaccurate, final text inaccurate	1.29	12.99
Improved		
Yes (improvement) source inaccurate, feedback accurate, final text accurate		87.01

Table 4. Revision types

Revision Type	Percentage
Macro-level meaning change (new content)	11.69
Macro-level meaning change (organization)	2.59
Micro-level meaning change (extended content)	5.19
Surface (formal)	28.59
Surface (meaning-preserving)	51.94

There are not a lot of feedbacks being given, especially regarding the overall structure of the text. It would be good if some feedback regarding the flow or effect of the translation when read can be commented. Most of the feedback revolves around suggestions for specific words or phrases.

With only 4.62 per cent of the total distribution of feedback types, questions are rare in the data set. The majority of the questions present are suggestions phrased as a question, e.g. “Which microstrategy is this?” and “Maybe x is better” and often contain a solution or hint at a possible solution using a politeness marker. Many of the statements (17.13 per cent of the total feedback types) have the same function, but the statements often co-occur with an explanatory phrase or sentence, e.g. “This link between macro and micro is missing in the others. It is a perfect transition”.

In the focus group interviews, students discussed feedback types and the majority agreed that isolated statements without further explanation are not particularly helpful for feedback implementation. Regarding feedback types provided, one student elaborated:

Student Peer Feedback in a Translation Task

They specifically mentioned what was wrong and it was explain why. Afterwards they suggested how they think it should change, even though they didn't impose their idea, they just made an example to make it easier for me to understand, which worked.

From the revisions made in their submitted assignments, a total of 216 implementation types were grouped into implemented feedback comments and non-implemented feedback comments (see Tables 2 and 3). These include a special case of one student who, instead of revising the assignment, provided a reflection for each feedback element judged worth implementing. The 133 instances of the non-implemented feedback comments were qualitatively examined to interpret why students might not have used the feedback. The 18.79 per cent non-implemented feedback comments allow assumptions about non-implementation – drawing on the additional open-ended questionnaire answers provided and based on statements from the focus group interviews. For the remaining high number of non-implemented comments (81.21 per cent), qualitative interpretation proves difficult, as these cases are not explicitly related to other data sets of the study. In the focus group interviews, one student states regarding non-implementation: *“You can always disagree”* and it is elaborated in the focus group discussion that the author has the authority and, in the end, *“the author decides what gets implemented”*.

Among the non-implemented feedback comments, the interpretable 18.79 per cent implementation types fall into the following categories: Comments that the peer clearly did not understand (thirteen instances), comments that were ignored because the feedback was obviously wrong (four instances), feedback comments that remained in the final version (three instances), feedback that consisted of comparisons with the feedback providers' own work, which the receiver knew and disagreed with (two instances), feedback on the translation that contradicted the original authors' commentary (two instances) and feedback that was related to another comment elsewhere (one instance) – with the problem solved elsewhere in the commentary.

A total of 77 feedback comments were implemented. These are represented in four different implementation categories, see Table 3. The implementation categories are based on teacher judgement³ of the outcome of the implementation.

Regarding teacher judgement, accuracy in the coding (as in “final text accurate”/“final text inaccurate”) did not necessarily mean an “adequate” translation solution (in the sense of faithfulness to the source text). Therefore, this study understands accurate/inaccurate as satisfactory/unsatisfactory or acceptable/unacceptable respectively and understands “accurate” to be a comprehensible English language expression in the given utterance (at sentence level). Yet, this coding was chosen because of the international students' Google Translation results and because only few students linked their peers' translation commentary with their feedback on both the translation and the translation commentary. Another reason is that seven of nineteen students received no feedback regarding their translation commentary and three students received only one comment each.

The implemented comments (12.99 per cent) which resulted in incorrect implementations fall into three subcategories. One student implemented four identical comments by three peers in the form of a descriptive note that explained the student's approach to solve the assignment. The descriptive note was clearly intended for the teacher. Five correct suggestions by peers were implemented but resulted in inaccurate solutions. One incorrect feedback comment led to a different, yet also incorrect implementation.

A majority of all implemented comments (87.01 per cent) resulted in a correct change that improved text quality. In the focus group interview comments and in the open-ended elaboration comments in the

questionnaire, students' reason for their implementations. The following example is from the open-ended questions section of the questionnaire – and one of many phrased similarly: *“The reason for implementing the feedback was good argumentation and obvious mistakes from my side, such as misspelling and lacking apostrophes. All the feedback which I received was implemented because I agree with their changes”*. The data from the special case of a student who provided a reflection for each feedback element is included in the data set (Table 4). In the special case, two out of eight comments contain vague suggestions, which fail to provide clear solutions/feed forward. Still, the receiver formulates an appreciative implementation comment that reflects superiority over the peer. In the focus group interview, students discussed their unequal status as learners and *“hierarchy”* in the groups. They stress that they typically understand what their peers mean with their comments and that *“all feedback is good feedback”* in the sense that also unclear feedback prompts their attention *“you are forced to look twice”* and makes them re-read and re-consider the section highlighted by the peer. In the words of the special case student, *“It is nice to see that people have another point of view than me – so you can take this up for discussion with yourself”*.

The revision type analysis qualifies the implementations further to distinguish between surface changes and micro-level and macro-level meaning changes.

Given the shortness of the source and target texts of the localization assignment and the limited number of translation commentary feedback comments, the high percentage of surface changes and the low number of macro-level meaning changes is not surprising. Yet, the few meaning changes implemented had a considerably positive impact on translation and commentary quality. The students themselves are aware of both the quantity and quality of the feedback provided and their relation to the final text product, as is expressed in this feedback process comment from a written questionnaire answer:

As the feedback from A was not that thorough it only helped my text a bit. A lot of the feedback from B was missing in the translation and that obviously made it difficult to improve my text. The feedback from C was good but I missed some comments on microstrategies.

Questionnaire Results

This section provides the results of the questionnaire (seven students' answers in writing). From the seven students who filled in the complete questionnaire in writing (and did not participate in the focus group interviews which used the same questions), all but one student answered all questions. The comments were comprehensive, ranging from 16-134 words. Regarding receiving feedback, the majority of both groups, namely seventeen students strongly agreed to *“I understand and know how to implement the feedback that I have received”*, one student is indifferent to this statement. This student received most feedback (32 comments) of which only two changes were implemented. Four of the feedback comments referred to a misleading approach in the student's translation commentary and were covered by a note to the teacher by the feedback receiver explaining the approach criticised by the peers. The note is one of the few HOCs in the data set, a macro-level meaning change regarding organisation. The other change implemented is based on a complex feed forward comment by one peer, a statement with a solution regarding a misinterpretation of translation microstrategies.

From the combined questionnaire and focus group interview answers, seventeen students agreed that the feedback that they had received *“... clearly told [them] what [they were] supposed to do to improve the text”*. The high number of correctly implemented feedback comments (87.01 per cent) from the implementation data analysis confirmed this self-evaluation. Twelve students agreed that the feedback

Student Peer Feedback in a Translation Task

was “sufficient and will help to improve my final text”. The implementation analysis confirmed this, yet when looking at the revision type analysis, the comments were mainly surface comments (LOCs), although 51.94 per cent were meaning-preserving, a category which exceeds mere spelling, tense and punctuation mistakes. Sixteen students agreed that “the feedback I received made me think about the point(s) indicated in the feedback and the task as a whole”. This was supported by positive elaboration statements in the open-ended question section of the questionnaire about the feedback approach. One elaboration comment particularly reflected this: *“The feedback made it easy to understand the specific task, but it also made sense regarding web-based communication in general and its terms and models”*. The majority of students agreed peer feedback to be “a worthwhile exercise in order to improve my writing and translation skills” (eighteen students) and to “like feedback from peers on my writing tasks” (thirteen students). This was confirmed by a control statement from the questionnaire form: “It is disheartening to receive negative feedback from my peers”, with which the majority disagreed (sixteen students). The overall positive reaction reflects the previously mentioned easy-to fix surface comments (LOCs) and the high number of motivational feedback types. Interestingly, fourteen students answered that they found “Feedback from my teacher is more useful than feedback from my peers” by ticking the box “agree” in the questionnaire. Only one student strongly agreed. A comment regarding feedback quality among peers as opposed to teacher feedback was found in an elaboration statement in the open-ended question section of the questionnaire: *“I hope that my comments and my feedback has helped the quality of their final texts. But I find it hard to know for sure as we are all just students, and their work might be better without the implementing of my suggestions”*. In both questionnaires and focus group interviews, unsurprisingly, the international students in particular voiced a preference for teacher feedback. In the questionnaires, ten students agreed to have provided sufficient feedback that “will help him/her to improve his/her revised text”. Seven students were indifferent to this statement, two disagreed. The statement whether the feedback provided was “likely to make him/her think about the point(s) indicated in my feedback and the task as a whole” was strongly agreed by two students, agreed by seven and eight students were indifferent. The feedback analysis data shows that the two students who valued their feedback highly also provided the most comprehensive feedback of the set. An elaboration statement in the open-ended question section of the questionnaire displays critical reflection in this regard: *“I am not sure if the peers are considering my feedback, one can only hope so. It depends on the peers own ideas are corresponding to those of mine”*. This statement does not only point to the feedback loop and the relevance of feedback-on-feedback, but also to issues of responsibility and authority. The latter were also discussed in the focus group interviews. Among the general statements about feedback, seventeen students considered providing peer feedback “a worthwhile exercise in order to improve my writing and translation skills”. This positive outcome of the study highlights the added value of peer feedback in translation teaching. In contrast, the evaluation statement, whether engaging in the feedback task “was a hassle-free and enjoyable experience” provided mixed results. The statement *“I like providing feedback to peers on their writing tasks”* also provided mixed results, with seven students who were indifferent to the statement and eight agreed. Interestingly, ten students chose the indifferent middle line to comment on the statement “I consider myself good at providing feedback”, six agreed to this, one person strongly agreed. A similar picture shows for the statement “I always consider the person I am giving feedback to and adapt my feedback to that person”, with six students agreeing and six students disagreeing. Three statements regarding the learning effect of peer feedback tasks were agreed to by the majority of students. The first stated “The peer feedback task allowed me to improve my final text considerably”, with four students strongly agreeing and eleven agreeing. The second statement “Reading my peers’ texts and

providing feedback has made me more aware of the weaknesses in my own writing”, is agreed by eight and strongly agreed to by four students. The third statement “Peer feedback tasks are well integrated into the web-based communication course” is strongly agreed to by four students and agreed to by twelve students. Students’ elaboration comments regarding far-reaching consequences of feedback practices applied in the course are positive: *“The feedback I gave was an indication of the problems they had with using references after statements in their commentary. This would give them the opportunity to remember to do this during exams and in other written work”*. Fourteen students agreed that they “understand and know how to provide feedback in different ways to ensure the feedback that others receive from me is understood”. The statement “I understood the task description and knew how to provide feedback to others” was also mainly agreed by fourteen students, four were indifferent and one student disagreed. Students’ knowledgeable ability about feedback processes and feedback types and their use and application of feedback terminology across the study confirms their general understanding of the concept.

Focus Group Interview Results

The focus group interviews comprised two interview groups with 5 and 7 students each, who elaborated on a number of points derived from the prompts. The resulting focus points in the focus group interviews share a number of similarities with the questionnaire group elaborations. In both groups, the quality of good argumentation for feedback solutions was evaluated highly, with one student stating that his peers *“must persuade me and work a bit hard”*, in order to convince him/her to implement the feedback. In this regard, the importance of strong claims by the feedback provider including argumentation are highlighted: *“claims without backing are not good”*. All students from the questionnaire group explicitly stated which types of feedback they had used and self-critical remarks were made regarding the quality of the comments in terms of use of course literature *“I think I used a lot of directives but also made some proposals. I did not quote the literature but tried to argument with it in my comments”*. Both groups highlighted that some but not all feedback was implemented. One focus group discussed implementation choices in the light of author responsibility and the perceived status of the group member providing feedback *“It depends who the feedback comes from ... I consider feedback much more when it comes from an authority”*. Both questionnaire and focus groups pointed out the difference between student and teacher feedback as in *“we are all students and might not have a 100% correct answer”*. One student described that teachers (also) often do not explicitly *“state what the problem in an assignment is in their feedback”*. But it was highlighted in the questionnaire by one student that *“it may be worth it to get more feedback from the teachers since they are the ones that know more about the topic and also one of them is a native English speaker”*. In terms of feedback types, the quality and the number of feedback comments received, the groups’ comments differ slightly. An overabundance of LOCs was commented on by the questionnaire group. One focus group discussed feedback types and errors made and the phrasing of feedback, with one student stating that introduction phrases like *“I would suggest...”* and *“Maybe you could...”* *“makes the reader more open to take in the feedback”*. This group agreed that blatant mistakes can be pointed out without explanations: *“I think with spelling mistakes, you should not make a full sentence”*. The questionnaire group and the focus groups stated uncertainty regarding the quality of the feedback they provided. Also, the issue of a Danish source text to be translated with Google Translator was raised by the international students *“I might not have gotten all the phrases right as I am not a Danish speaker”* and one focus group discussed the superiority of the Danish students *“their level of English is better than mine”*. In this regard, the international students mentioned the L2 language barrier

Student Peer Feedback in a Translation Task

and benefits from feedback on linguistic issues and content comments regarding translation strategies by their Danish peers, which they have *“of course implemented because it made my text better”*. Likewise, in the focus group interview, the Danish students stressed that they exerted care to provide more detailed comments to the international students. Time frame issues were mentioned, one student pointing out that providing feedback for an assignment that has not been entirely proofread by the author can cause a problem for the feedback provider: *“Providing peer-feedback on a task which was rushed, can cause some complications when providing feedback, because many things which have been commented on, such as the commentary and the small mistakes may not have been there if there had been more time”*.

In addition to the above, the focus group interview exploited a number of aspects that were not touched upon by the questionnaire group. The responsibility issue mentioned above led one focus group to a discussion about internal group hierarchies, where they compared student abilities, study years and language capacity (L2), with the international students stating about their counterparts, that *“The Danish people have a better level of English”*. The group agreed in the discussion that the teams know each other’s strengths and weaknesses. A Danish student stressed not to have relied on an international student’s correction of typos but that content comments by that peer were *“highly appreciated”*. One focus group also discussed the phrasing of feedback in relation with supportive, positive, encouraging comments. All but one student in the group agreed that motivational comments were important for their implementation processes and for feedback tasks in general and that these are enjoyed: *“I like giving positive feedback”*. One group discussed the time issue with task completion *“good phrasing needs time”*, in relation with the number of comments provided and received. Some students were apparently uncertain as to the error categories they needed to focus on and demanded more scaffolding for error types and feedback types in future feedback tasks. These uncertainties can be solved in future feedback tasks by specifying feedback types or e.g. a minimum or maximum number range of error or feedback types. Both focus groups also pointed out that there is a difference between written and oral feedback and a preference for dialogue. *“I prefer to actually to talk to the person I have to give feedback to, to know why did he do that and then have a direct conversation about how to change it maybe or maybe he has a good point and then I don’t even need to comment on it. That is why I prefer verbal conversation about feedback”*. Other students suggested implementation of both written and oral feedback in future peer feedback tasks. In the oral versus written feedback discussion, concerns were raised that the oral feedback *“is too spontaneous”* but also *“more convincing”* because of the face-to-face contact. Concerns about losing face in oral feedback *“written feedback is more anonymous”* were also raised. Asked about personal strengths in providing feedback and focus points, students pointed out understandability, academic guidelines and course content. A consensus was found that multiple peer feedback *“is a good technique”* that *“should become a student habit”*. The students agreed that it was good that they were *“forced to give some kind of criticism”* and *“to look again at your writing”* and stressed the fact that the feedback practices learned in the course were transferable and applicable to other courses and writing tasks. Another consensus reaching across the two focus interview groups was the importance of good argumentation on the part of the feedback provider. The social dimension was highlighted and a student pointed out that she was delighted that someone else but the teacher *“cares about my work”*. Another aspect relating to the social dimension, group composition, brought to light that groups knowing each other operate differently, supposedly using an explicit or implicit code in their feedback, as they know (or believe to know) what the other group members make of their comments. One student pointed out a learning effect in dealing with feedback by comparing the feedback comments for an earlier feedback task in the semester and the degree of detail applied and the helpfulness of the feedback task from this

study. One of the focus group interview participants pointed out that the multiple peer feedback task allowed for a comparison of their work and their feedback with others: “*you get to see how others do it*” and that “*receiving feedback has an impact on providing feedback*”.

Key Findings and Limitations of the Study

Training in peer feedback during the course ensured that students were able to perform to the teacher’s satisfaction in the multiple peer feedback task. Drawing on feedback literature scaffolding was applied (Min, 2005; Van Steendam et al., 2014) and it worked, besides the limitation that the task description neither identified learning objectives nor highlighted feedback areas or specified success criteria to focus on, as recent feedback literature suggests (Nicol, 2011; Nicol et al., 2014). This might have led to students’ underestimation of the overlap between web-localization task and commentary, resulting in only few macro-level meaning change comments (HOCs). This result is not only in line with feedback research in areas other than localisation and translation commentary writing (Tsui & Ng, 2000; Hattie & Timperley, 2007), it also bears in itself the requirement for future feedback implementation training. Still, a comprehensive range of feedback and implementation types was represented in the data, not unexpectedly, as feedback types had been introduced during the course to ensure targeted student feedback. Instructions were given to encourage the students to perform beyond their current level of performance, an aspect confirmed by the data. Compared with a previous study in the same course (Flanagan & Heine, 2015), considerably fewer student feedback comments contained vague questions or statements without problems solving, an aspect that supports the notion that scaffolding peer feedback has its place in teaching approaches. The exact degree to which the scaffolding has influenced the feedback types cannot be provided, but the approach of the present study lends itself to further investigations with a pre-test/post-test design and even more detailed scaffolding. Such scaffolding could include references to professional practices, to make the students aware that peer feedback is not only a didactic tool, but a skill required in the profession. The majority of the students perceived the multiple peer feedback task as a worthwhile experience. Proficiency level comments on analytic ability, grammar, knowledge of metalanguage and appropriate use of course terminology trigger the authors’ sensitivity for pair composition and roles (as discussed in Davidson, 1994) and skill diversity in future feedback assignments. Likewise, the study draws attention to future incorporation of focus points for feedback into instructions, e.g. linguistic aspects, course content and academic writing features, in other words, extended scaffolding. The combination of questionnaire and focus group interview follow-up raised teacher and student awareness of the need for student dialogue about feedback before and after implementation and teacher feedback-on-feedback, especially since the data shows student difficulties in evaluating the feedback they provide. Drawing on Nicol (2010), future task descriptions of peer feedback assignments should include dialogue as a set element of the feedback process. Regarding feedback-on-feedback, teacher advice is clearly missing at this under-researched end of the feedback loop. While students apparently have a clear understanding of how to deal with the feedback they receive, the uncertainty how their feedback is understood and whether it gets implemented by the peer, leaves feedback providers clueless as to their comments’ impact and relevance for the feedback receiver. This highlights a limitation of this study, as a dialogical component before or upon implementation was not incorporated. Active and reflective (student-student) interpretation tasks would encourage students to compare feedback against the expected performance to derive implications for further learning.

CONCLUSION

Students performed well in this multiple peer feedback research project, given the limitations of the scaffolding approach applied. Given the small scale of the study, generalizations cannot be drawn. Yet, this study contributes to the greater body of knowledge about peer feedback by highlighting the importance of raising student awareness to task, feedback and implementation requirements and to the importance of task reflection. To this author's knowledge, implementation requirements and task reflection have not been discussed in any detail in the literature on peer feedback in translation didactics. In the future, this gap could possibly be filled by feedback training, dialogue peer feedback and feedback-on-feedback provided by the teachers, designed to match assessment criteria of the course and professional practice. Such a feedback setup requires further in-depth scholarly attention, as it appears to be an under-researched area – not only in Translation Studies. Information about feedback and quality assurance practices from the profession should be incorporated into future peer feedback practice. The questionnaire used provided reliable and accurate data. Focus group interviews found their place in this study, as they contributed substantially more meaningful data than the questionnaire data alone would have provided. “Odd ones out” questionnaire numbers could be explained drawing on the focus group interviews, providing valuable insights into the individuals' perceptions of feedback that would have gone unnoticed with qualitative feedback data analysis alone. Student elaborations on the interview topics were priceless sources of information. As relevant as numerical proof might be to illustrate the use of feedback, revision and implementation types, from a didactics viewpoint, participant comments on the above exceeds it and provides further insight. Many data have not been discussed yet: aspects like sense of ownership, sufficiency of own engagement and other-engagement, time investment, composition, multi-facetedness of the task, group composition and skill diversity, authorship, argumentation, experiences across the boundaries of the task, and the relevance of L2-language proficiency levels (of the heterogeneous group) were all raised by the students. These issues require further scholarly attention. A focus on feedback training to help developing students' capacity to make qualitative judgements – as will be required from them in the profession – belongs on the Translation Didactics agenda. Likewise, feedback-on-feedback, implementation time, incorporation of assessment criteria, role descriptions incorporated into the classroom practice – translated into longitudinal, (statistical) research and pre-test/post-test designs to delve into these – are aspects for future research into peer feedback in Translation Didactics.

ACKNOWLEDGMENT

The author is gratefully indebted to co-teacher and partner in crime in peer feedback research, Marian Flanagan, for her invaluable support in the planning and realisation of this research project and for fruitful discussions on the way.

REFERENCES

- Barker, M., & Pinard, M. (2014). Closing the feedback loop? Iterative feedback between tutor and student in coursework assessments. *Assessment & Evaluation in Higher Education*, 39(8), 899–915. doi:10.1080/02602938.2013.875985
- Blair, A., Curtis, S., Goodwin, M., & Shields, S. (2013). What feedback do students want? *Politics*, 33(1), 66–79. doi:10.1111/j.1467-9256.2012.01446.x
- Boud, D. (2007). Reframing assessment as if learning were important. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education* (pp. 14–26). London: Taylor & Francis.
- Caulk, N. (1994). Comparing teacher and student responses to written work. *TESOL Quarterly*, 28(1), 181–188. doi:10.2307/3587209
- Chang, C. Y. (2016). Two decades of research in L2 peer review. *Journal of Writing Research*, 8(1), 81–117. doi:10.17239/jowr-2016.08.01.03
- Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction*, 20(4), 328–338. doi:10.1016/j.learninstruc.2009.08.006
- Cho, Y. H., & Cho, K. (2011). Peer reviewers learn from giving comments. *An International Journal of the Learning Sciences*, 39(5), 629.
- Davidson, N. (1994). Cooperative and collaborative learning. In J. S. Thousand, R. A. Villa, & A. Nevin (Eds.), *Creative and collaborative learning* (pp. 13–30). Baltimore, MD: Paul H. Brookes.
- Earley, P. C., Northcraft, G. B., Lee, C., & Lituchy, T. R. (1990). Impact of process and outcome feedback on the relation of goal setting to task performance. *Academy of Management Journal*, 33(1), 87–105. doi:10.2307/256353
- Faigley, L., & Witte, S. (1981). Analyzing revision. *College Composition and Communication*, 32(4), 400–414. doi:10.2307/356602
- Falchikov, N. (2005). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. London: Routledge-Falmer.
- Fitzgerald, J. (1992). *Knowledge in writing: Illustration from revision studies*. New York: Springer. doi:10.1007/978-1-4612-2796-0
- Flanagan, M., & Heine, C. (2015). Peer-feedback as a translation training tool in web-based communication. *Hermes*, 54, 115–136.
- Hattie, J., Gan, M., & Brooks, C. (2011). Instruction based on feedback. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 290–324). New York: Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. doi:10.3102/003465430298487

Student Peer Feedback in a Translation Task

- Hounsell, D. (1997/2005). Contrasting conceptions of essay-writing. In F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The Experience of learning: Implications for teaching and studying in higher education* (pp. 106–125). Edinburgh, UK: University of Edinburgh, Centre for Teaching, Learning and Assessment.
- Huertas Barros, E. (2011). Collaborative learning in the translation classroom: Preliminary survey results. *The Journal of Specialised Translation*, 16, 42–60.
- Krest, M. (1988). Monitoring student writing: How not to avoid the draft. *Journal of Teaching Writing*, 7(1), 27–39.
- Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research*, 47(2), 211–232. doi:10.3102/00346543047002211
- Lai, Y. (2010). Which do students prefer to evaluate their essays: Peers or computer program. *British Journal of Educational Technology*, 41(3), 432–454. doi:10.1111/j.1467-8535.2009.00959.x
- Leijen, D. (2014). Applying machine learning techniques to investigate the influence of peer feedback on the writing process. In D. Knorr, C. Heine, & J. Engberg (Eds.), *Methods in writing process research* (pp. 167–183). Frankfurt am Main: Peter Lang.
- Lindgren, E., Sullivan, K. P. H., Deutschmann, M., & Steinvall, A. (2009). Supporting learner reflection in the language translation class. *International Journal of Information Communication Technologies and Human Development*, 1(3), 26–48. doi:10.4018/jictd.2009070102
- Lisaitė, D., Maylath, B., Moustén, B., Valdez, S., Castel-Branco, M., Minacori, P., & Vandepitte, S. (2016). Negotiating meaning at a distance: Peer feedback in electronic learning translation environments in the Trans-Atlantic and Pacific Project. In M. Thelen & B. Lewandowska-Tomaszczyk (Eds.), *Translation and meaning. Part 10. Proceedings of the Maastricht Session of the 6th International Maastricht-Łódź Duo Colloquium on “Translation and Meaning”, 21-22 May 2015* (pp. 99-114). Maastricht: Zuyd University, Maastricht School of International Communication, Department of Translation and Interpreting.
- Macfarlane-Dick, D., & Nicol, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. doi:10.1080/03075070600572090
- Mangelsdorf, K., & Schlumberger, A. (1992). ESL student response stances in a peer-review task. *Journal of Second Language Writing*, 1(3), 235–254. doi:10.1016/1060-3743(92)90005-A
- McMillan, K., & Weyers, J. (2011). *How to write dissertations and project reports* (2nd ed.). Harlow, UK: Pearson Education.
- Min, H.-T. (2005). Training students to become successful peer reviewers. *System*, 33(2), 293–308. doi:10.1016/j.system.2004.11.003
- Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science*, 37(4), 375–401. doi:10.1007/11251-008-9053-x
- Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*, 35(5), 501–517. doi:10.1080/02602931003786559

- Nicol, D. (2011). *Developing the students' ability to construct feedback*. Gloucester, UK: Quality Assurance Agency for Higher Education.
- Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education*, 39(1), 102–122. doi:10.1080/02602938.2013.795518
- Price, M., & O'Donovan, B. (2006). Improving student performance through enhanced student understanding of criteria and feedback. In C. Bryan & K. Clegg (Eds.), *Innovative assessment in higher education* (pp. 100–109). London: Routledge.
- Puntambekar, S., & Hubscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist*, 40(1), 1–12. doi:10.120715326985ep4001_1
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144. doi:10.1007/BF00117714
- Schiefele, U. (1991). Interest, learning, and motivation. *Educational Psychologist*, 26(3-4), 299–323. doi:10.1080/00461520.1991.9653136
- Shei, C. C. C. (2005). Translation commentary: A happy medium between translation curriculum and EAP. *System*, 33(2), 309–325. doi:10.1016/j.system.2005.01.004
- Strijbos, J.-W., Narciss, S., & Dünneber, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction*, 20(4), 291–303. doi:10.1016/j.learninstruc.2009.08.008
- Thompson, I. (2009). Scaffolding in the writing center: A microanalysis of an experienced tutor's verbal and nonverbal tutoring strategies. *Written Communication*, 26(4), 417–453. doi:10.1177/0741088309342364
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249–276. doi:10.3102/00346543068003249
- Topping, K. (2009). Peer assessment. *Theory into Practice*, 48(1), 20–27. doi:10.1080/00405840802577569
- Tsui, A. B. M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing*, 9(2), 147–170. doi:10.1016/S1060-3743(00)00022-9
- Van Steendam, E., Rijlaarsdam, G. C. W., Van den Bergh, H. H., & Sercu, L. (2014). The mediating effect of instruction on pair composition in L2 revision and writing. *An International Journal of the Learning Sciences*, 42(6), 905–927.
- Vandepitte, S. (2016). Collaboration in translation practices: A course design for peer feedback in translation training. *Konin Language Studies*, 4(4), 289–304.
- Wang, K., & Han, C. (2013). Accomplishment in the multitude of counsellors: Peer feedback in translation training. *Translation & Interpreting*, 5(2), 62–75.
- Wang, W. (2014). Students' perceptions of rubric-referenced peer feedback on EFL writing: A longitudinal inquiry. *Assessing Writing*, 19, 80–96. doi:10.1016/j.asw.2013.11.008

ENDNOTES

- ¹ The course was co-taught by fellow researcher Marian Flanagan, who also participated in method design and focus group interviews.
- ² The study was carried out in 2016, when feedback was prominently discussed as a crucial element of learning at Danish Universities, while a country-wide study initiated by the Ministry of Education indicated that students at AU were not entirely satisfied with the feedback they received from both peers and members of faculty during their BA and MA courses – if they received any at all.
- ³ This researcher coded the data based on a coding scheme previously used in Flanagan and Heine (2015).

Compilation of References

- Abdallah, K. (2012). *Translators in production networks. Reflections on agency, quality and ethics*. Joensuu, Finland: Publications of the University of Eastern Finland.
- Adab, B. J. (2000). Evaluating translation competence. In C. Schäffner & B. J. Adab (Eds.), *Developing translation competence* (pp. 215–228). Amsterdam: John Benjamins. doi:10.1075/btl.38.20ada
- ADLAB PRO. (2017a). *Report on IO1: Assessment of current AD training practices*. Retrieved from <http://www.adlabproject.eu>
- ADLAB PRO. (2017b). *Report on IO2: Audio description professional: Profile definition*. Retrieved from <http://www.adlabproject.eu>
- ADLAB PRO. (n.d.). Retrieved from <https://adlabpro.wordpress.com>
- Agost, R. (1999). *Traducción y doblaje: Palabras, voces e imágenes*. Barcelona, Spain: Ariel.
- Alfaro, C. (2012). Quality standards or censorship? Language control policies in cable TV subtitles in Brazil. *Meta: journal des traducteurs/Meta: Translators' Journal*, 57(2), 464–477.
- Alonso Bacigalupe, L. (2013). Interpretation quality: From cognitive constraints to market limitations. In R. Barranco-Droegge, E. Macarena Pradas Macías, & O. García Becerra (Eds.), *Quality in interpreting: Widening the scope* (Vol. 2, pp. 9–33). Granada: Comares.
- Al-Qinai, J. (2000). Translation quality assessment: Strategies, parameters and procedures. *Meta*, 45(3), 497–519. Retrieved July 2, 2017, from <https://www.erudit.org/fr/revues/meta/2000-v45-n3-meta162/001878ar.pdf>
- ALTA Language Services, Inc. (2010). *Beyond words: California assessment of the Consortium for Language Access in the Courts' exams*. Retrieved from <http://www.courts.ca.gov/documents/ALTAReport.pdf>
- Alves, F. (2005). Bridging the Gap Between Declarative and Procedural Knowledge in the Training of Translators: Meta-Reflection Under Scrutiny. *Meta*, 50(4). Retrieved from <http://www.erudit.org/revue/meta/2005/v50/n4/019861ar.pdf>
- Alves, F. (Ed.). (2003). *Triangulating Translation: Perspectives in Process Oriented Research*. Amsterdam: John Benjamins. doi:10.1075/btl.45
- Alves, F., & Hurtado-Albir, A. (2017). Perspectives for research on cognitive aspects of translation. In J. W. Schwieter & A. Ferreira (Eds.), *The handbook of translation and cognition* (pp. 537–553). Oxford, UK: Wiley Blackwell. doi:10.1002/9781119241485.ch29
- Alves, F., & Vale, D. C. (2011). On drafting and revision in translation: A corpus linguistics oriented analysis of translation process data. *Translation: Corpora, Computation, Cognition. Special Issue on Parallel Corpora: Annotation, Exploitation, Evaluation*, 1(1), 105–122.

Compilation of References

- Anckaert, P., Eyckmans, J., & Segers, W. (2006). Vertaalvaardigheid evalueren: Een normgerelateerde benadering. *Tijdschrift nff*, 6, 9–27.
- Anckaert, P., & Eyckmans, J. (2007). IJken en ijken is twee. Naar een normgerelateerde ijkpuntenmethode om vertaalvaardigheid te evalueren. In C. Van de Poel & W. Segers (Eds.), *Vertalingen objectief evalueren: Matrices en Ijkpunten* (pp. 53–81). Leuven: Acco.
- Anckaert, P., Eyckmans, J., & Segers, W. (2008). Pour une évaluation normative de la compétence de traduction. *ITL: Review of Applied Linguistics*, 41(155), 53–76. doi:10.2143/ITL.155.0.2032360
- Angelelli, C. (2006). Validating professional standards and codes: Challenges and opportunities. *Interpreting*, 8(2), 175–193. doi:10.1075/intp.8.2.04ang
- Angelelli, C. V. (2009). Using a rubric to assess translation ability: Defining the construct. In C. V. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in Translation and Interpreting Studies: A call for dialogue between research and practice* (pp. 13–47). Amsterdam: John Benjamins. doi:10.1075/ata.xiv.03ang
- Angelelli, C. V., & Jacobson, H. E. (Eds.). (2009). *Testing and assessment in translation and interpreting studies. A call for dialogue between research and practice*. Amsterdam: John Benjamins.
- Angelelli, C., & Jacobson, H. E. (Eds.). (2009). *Testing and assessment in translation and interpreting studies*. Amsterdam: John Benjamins. doi:10.1075/ata.xiv
- Angelone, E. (2010). Uncertainty, Uncertainty Management, and Metacognitive Problem Solving in the Translation Task. In G. Shreve & E. Angelone (Eds.), *Translation and Cognition* (pp. 17–40). Amsterdam: John Benjamins. doi:10.1075/ata.xv.03ang
- Angelone, E. (2012). The Place of Screen Recording in Process-Oriented Translator Training. *RITT*, 14, 41–55.
- Angelone, E. (2013). The impact of process protocol self-analysis on errors in the translation product. *Translation and Interpreting Studies*, 8(2), 253–271. doi:10.1075/tis.8.2.07ang
- Angelone, E. (2014). A Corpus-Based Comparison of Self-Reflection Modalities in Process-Oriented Translator Training. In Y. Cui & W. Zhao (Eds.), *Teaching Language Translation and Interpretation: Methods, Theories, and Trends* (pp. 346–361). Hershey, PA: IGI Global.
- Angelone, E. (2015). Optimizing Process-oriented Translator Training Using Freeware. In P. Sandrini & M. García González (Eds.), *Translation and Openness* (pp. 133–145). Innsbruck: University of Innsbruck Press. doi:10.15203/2936-88-2-7
- Angelone, E. (2016). A process-oriented approach for documenting and gauging intercultural competence in translation. *The Interpreter and Translator Trainer*, 10(3), 304–317. doi:10.1080/1750399X.2016.1236560
- Angelone, E., & Shreve, G. (2011). Uncertainty Management, Metacognitive Bundling in Problem-Solving, and Translation Quality. In S. O'Brien (Ed.), *Cognitive Explorations of Translation* (pp. 108–130). London: Continuum.
- Antonini, R. (2005). The perception of subtitled humour in Italy. *Humor: International Journal of Humor Research*, 18(2), 209–225. doi:10.1515/humr.2005.18.2.209
- Argenti, P. A. (2013). *Corporate communication* (6th ed.). New York: McGraw-Hill Education.
- Arjona, E. (1985). The court interpreters certification test design. In L. Elías-Olivares, E. A. Leone, R. Cisneros, & J. R. Gutiérrez (Eds.), *Spanish language use and public life in the USA* (pp. 181–200). Berlin: Mouton de Gruyter.

- Arjona-Tseng, E. (1994). A psychometric approach to the selection of translation and interpreting students in Taiwan. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation* (pp. 69–85). Amsterdam: John Benjamins. doi:10.1075/btl.3.08arj
- Arumí Ribas, M. (2010). Formative assessment in the interpreting classroom: Using the portfolio with students beginning simultaneous interpreting. *Current Trends in Translation Teaching and Learning E*. Retrieved from <http://www.cttl.org/uploads/5/2/4/3/5243866/currenttrendstranslation2010.pdf>
- Arumí Ribas, M. (2010). Formative assessment in the interpreting classroom: Using the portfolio with students beginning simultaneous interpreting. *Current Trends in Translation Teaching and Learning*, 3, 97–133.
- Arumí, M., Carrabina, J., Matamala, A., Mesa-Lao, B., Orero, P., & Serrano, J. (2013). Audiovisual translation learning platform: A new tool for audiovisual translation training. *Hermeneus*, 15, 39–66.
- Arumí, M., & Esteve, O. (2006). Using instruments aimed at self-regulation in the consecutive interpreting classroom. *Electronic Journal of Foreign Language Teaching*, 3(2), 158–189.
- Assessment Standards Knowledge exchange (ASKe). (n.d.). *Assessment: An ASKe position paper*. Retrieved July 7, 2017, from <https://www.brookes.ac.uk/WorkArea/DownloadAsset.aspx?id=2147552601>
- ASTM International. (2006). *F2575-06 standard guide for quality assurance in translation*. West Conshohocken, PA: ASTM International.
- Audio Description Association. (2004). *Certificate in Audio Description Skills Training Course Manual* (unpublished).
- Babych, B., Hartley, A., Kageura, K., Thomas, M., & Utiyama, M. (2012). MNH-TT: A collaborative platform for translator training. *Translating and the Computer*, 34, 1–18.
- Baker, M. (1992). *In other words: A coursebook on translation*. London: Routledge. doi:10.4324/9780203327579
- Barker, M., & Pinard, M. (2014). Closing the feedback loop? Iterative feedback between tutor and student in coursework assessments. *Assessment & Evaluation in Higher Education*, 39(8), 899–915. doi:10.1080/02602938.2013.875985
- Bartoń, K. (2009). *MuMIn: Multi-model inference*. R Package Version 0.12.2/r18. Retrieved from <https://CRAN.R-project.org/package=MumIn>
- Bartrina, F. (2005). Theory and translation training. In M. Tennent (Ed.), *Training for the new millennium* (pp. 177–189). Amsterdam: John Benjamins. doi:10.1075/btl.60.15bar
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using {lme4}. *Journal of Statistical Software*, 67(1), 1–48. doi:10.18637/jss.v067.i01
- Baxter, R. N. (2016). Aproximación a los posibles efectos del uso de presentaciones visuales sobre la sincronía y el décalage en la interpretación simultánea. *Sendebarr*, 27, 9–23.
- Bayer-Hohenwarter, G. (2011). Kreativität in populärwissenschaftlichen und instruktiven Texten im Vergleich: Kein Raum für Kreativität beim Übersetzen von instruktiven Texten? *trans-kom*, 4(1), 49-75.
- Bayer-Hohenwarter, G. (2010). Comparing translational creativity scores of students and professionals: Flexible problem-solving and/or fluent routine behaviour? In S. Göpferich, F. Alves, & I. M. Mees (Eds.), *New approaches in translation process research* (pp. 83–111). Copenhagen: Samfundslitteratur Press.
- Beeby, A., & Martínez, N. (1998). *La evaluación en la enseñanza universitaria de la traducción profesional: Un estudio empírico*. Paper presented at the EST Congress, Granada.

Compilation of References

- Beeby, A. (2000). Evaluating the Development of Translation Competence. In C. Schäffner & B. Adab (Eds.), *Developing translation competence* (pp. 185–198). Amsterdam: John Benjamins. doi:10.1075/btl.38.18bee
- Beeby, A., Fernández, M., Fox, O., Albir, A., Kozlova, I., Kuznik, A., ... Wimmer, S. (2009). Results of the validation of the PACTE translation competence model: Acceptability and decision making. *Across Languages and Cultures*, 10(2), 207–230. doi:10.1556/Acr.10.2009.2.3
- Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education*, 85(5), 536–553. doi:10.1002/ce.1022
- Bell, R. T. (1991). *Translation and translating: Theory and practice*. New York: Longman London.
- Bell, R. T. (1998). Psycholinguistic/cognitive approaches. In M. Baker (Ed.), *Routledge encyclopedia of translation studies* (pp. 185–190). London: Routledge.
- Benecke, B. (2014). *Audiodeskription als partielle Translation. Modell und Methode*. Berlin: LitVerlag.
- Benson, P. (2001). *Teaching and researching autonomy in language learning*. London: Longman.
- Berg, E. C. (1999). The effects of trained peer response on ESL students' revision types and writing quality. *Journal of Second Language Writing*, 8(3), 215–241. doi:10.1016/S1060-3743(99)80115-5
- Biel, Ł. (2011). Professional realism in the legal translation classroom: Translation competence and translator competence. *Meta*, 56(1), 162–178. doi:10.7202/1003515ar
- Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). Berkshire, UK: Open University Press.
- Biocca, F., & Harms, C. (2002). Defining and measuring social presence: Contribution to the networked minds theory and measure. *Proceedings of PRESENCE, 2002*, 7–36.
- Bittner, H. (2011). The quality of translation in subtitling. *Trans-kom*, 4(1), 76–87.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. doi:10.1080/0969595980050102
- Blair, A., Curtis, S., Goodwin, M., & Shields, S. (2013). What feedback do students want? *Politics*, 33(1), 66–79. doi:10.1111/j.1467-9256.2012.01446.x
- Bloom, B. S. (1972). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive Domain*. New York: David McKay.
- Bloom, B. S., Hastings, J. T., & Madaus, G. F. (Eds.). (1971). *Handbook on the formative and evaluation of student learning*. New York: McGraw-Hill.
- Bloxham, S. (2015). Assessing assessment. New developments in assessment design, feedback practices and marking in higher education. In H. Fry, S. Ketteridge, & S. Marshall (Eds.), *A handbook for teaching and learning in higher education: Enhancing academic practice* (4th ed.; pp. 107–122). London: Routledge.
- Bontempo, K., & Napier, J. (2009). Getting it right from the start. Program admission testing of signed language interpreters. In C. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies* (pp. 246–295). Amsterdam: John Benjamins. doi:10.1075/ata.xiv.13bon
- Boud, D., & Associates. (2010). *Assessment 2020: Seven propositions for assessment reform in higher education*. Australian Learning and Teaching Council. University of Technology, Sydney. Retrieved July 7, 2017, from https://www.uts.edu.au/sites/default/files/Assessment-2020_propositions_final.pdf

- Boud, D. (1988). Moving towards autonomy. In D. Boud (Ed.), *Developing student autonomy in learning* (2nd ed.; pp. 17–39). London: Kogan Page.
- Boud, D. (1992). The use of self-assessment schedules in negotiated learning. *Studies in Higher Education*, 17(2), 185–200. doi:10.1080/03075079212331382657
- Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.
- Boud, D. (2007). Reframing assessment as if learning were important. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education* (pp. 14–26). London: Taylor & Francis.
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31(4), 399–413. doi:10.1080/02602930600679050
- Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education*, 38(6), 698–712. doi:10.1080/02602938.2012.691462
- Bourne, J., & Lachat, C. (2010). Impacto de la norma AENOR: Valoración del usuario. In C. Jiménez, A. Rodríguez, & C. Seibel (Eds.), *Un corpus de cine. Teoría y práctica de la audiodescripción* (pp. 315–333). Granada, Spain: Tragacanto.
- Bowker, L. (2001). Towards a methodology for a corpus-based approach to translation evaluation. *Meta: Journal des Traducteurs/Meta: Translators' Journal*, 46(2), 345–364.
- Braun, S. (2007). Audio description from a discourse perspective: A socially relevant framework for research and training. *Linguistica Antverpiensia, New Series. Themes in Translation Studies*, 6, 357–369.
- Brown, G., Bull, J., & Pendlebury, M. (1997). *Assessing student learning in higher education*. London: Routledge.
- Brown, S. (2015). *Learning, teaching and assessment in higher education: Global perspectives*. London: MacMillan Education. doi:10.1007/978-1-137-39667-9
- BS EN-15038. (2006). *European quality standard. European quality standard for translation service providers*. Brussels: European Committee for Standardization.
- Bucaria, C., & Chiaro, D. (2007). End-user perception of screen translation: The case of Italian dubbing. *TradTerm*, 13, 91–118. doi:10.11606/issn.2317-9511.tradterm.2007.47468
- Bühler, H. (1986). Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters. *Multilingua*, 5(4), 231–235.
- Cabeza Cáceres, C. (2017, March). *Professional audio descriptions: Same film, same assignment = same quality?* Paper presented to Advanced research seminar in audio description (ARSAD), Barcelona, Spain. Retrieved from <http://grupsderecerca.uab.cat/arsad/sites/grupsderecerca.uab.cat.arsad/files/cabeza-cacerescristobal-arsad2017.pdf>
- Caestecker, F. (2016). *Hoe de mens de wereld vorm gaf*. Ghent, Belgium: Academia Press.
- Cahill, L. (2002). Reflecting on peer review practices. In D. Roen, V. Pantoja, L. Yena, S. Miller, & E. Waggoner (Eds.), *Strategies for teaching first-year composition* (pp. 301–306). Urbana, IL: NCTE.
- Calvo, E. (2015). Scaffolding translation skills through situated training approaches: Progressive and reflective methods. *The Interpreter and Translator Trainer*, 9(3), 306–322. doi:10.1080/1750399X.2015.1103107
- Caminade, M., & Pym, A. (1998). Translator training institutions. In M. Baker (Ed.), *Routledge encyclopaedia of Translation Studies* (pp. 280–285). London: Routledge.

Compilation of References

- Campbell, S. (1991). Towards a model of translation competence. *Meta*, 36(2-3), 329–343. doi:10.7202/002190ar
- Campbell, S. (1998). *Translation into the second language*. London: Longman.
- Campoy-Cubillo, M. C. (2017). Listening skills: Issues in assessment and implementation. In V. Bonsignori & B. C. Camiciottoli (Eds.), *Multimodality across communicative settings, discourse domains and genres* (pp. 14–37). Newcastle-upon-Tyne, UK: Cambridge Scholars Publishing.
- Candy, P. (1988). On the attainment of subject-matter autonomy. In D. Boud (Ed.), *Developing student autonomy in learning* (2nd ed.; pp. 59–76). London: Kogan Page.
- Canfora, C., & Ottmann, A. (2015). Risikomanagement für Übersetzungen. *trans-kom*, 8(2), 314-346. Retrieved April 10, 2017, from http://www.trans-kom.eu/bd08nr02/trans-kom_08_02_02_Canfora_Ottmann_Risiko.20151211.pdf
- Canon, M., & Marifian, E. (2013). Job polarization leaves middle-skilled workers out in the cold. *The Regional Economist*, 9-11. Retrieved June 9, 2017, from https://www.stlouisfed.org/~media/Files/PDFs/publications/pub_assets/pdf/re/2013/a/employment.pdf
- Carl, M., Kay, M., & Jensen, K. T. H. (2010). Long distance revisions in drafting and post-editing. *Proceedings of CICLing-2010*. Retrieved from http://static-curis.ku.dk/portal/files/131448111/2010_Long_Distance_Revisions_CarlKayJensen.pdf
- Carl, M., Bangalore, S., & Schaeffer, M. (Eds.). (2016). *New directions in empirical translation process research: Exploring the CRITT TPR-DB*. Cham: Springer International. doi:10.1007/978-3-319-20358-4
- Carl, M., Gutermuth, S., & Hansen-Schirra, S. (2015). Post-editing machine translation: Efficiency, strategies, and revision processes in professional translation settings. In A. Ferreira & J. W. Schwieter (Eds.), *Psycholinguistic and cognitive inquiries into translation and interpreting* (pp. 145–174). Amsterdam: Benjamins. doi:10.1075/btl.115.07car
- Carl, M., & Kay, M. (2011). Gazing and typing activities during translation: A comparative study of translation units of professional and student translators. *Meta: Translators' Journal*, 56(4), 952–975.
- Carl, M., Schaeffer, M. J., & Bangalore, S. (2016). The CRITT translation process research database. In M. Carl, S. Bangalore, & M. Schaeffer (Eds.), *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB. New Frontiers in Translation Studies* (pp. 13–54). Cham: Springer International Publishing.
- Castillo Rincón, L. M. (2010). *La evaluación de la aceptabilidad de las traducciones: Un estudio exploratorio sobre la idoneidad de los puntos ricos para evaluar traducciones* (Master's thesis). Universitat Autònoma de Barcelona, Barcelona, Spain. Retrieved June 22, 2017, from <http://hdl.handle.net/2072/179692>
- Castillo Rincón, L. M. (2015). Acquisition of translation competence and translation acceptability: An experimental study. *The International Journal for Translation & Interpreting Research*, 7(1), 72–85.
- Catford, J. (1965). *A linguistic theory of translation*. Oxford, UK: Oxford University Press.
- Caulk, N. (1994). Comparing teacher and student responses to written work. *TESOL Quarterly*, 28(1), 181–188. doi:10.2307/3587209
- Cavallo, A. (2016). Audio description: Ramps on the moon. *Exeunt*. Retrieved from <http://exeuntmagazine.com/features/audio-description-ramps-moon/>
- CEN. (2006). *Translation services: Service requirements, EN 15038*. Brussels: CEN.
- Chandler, D., & Munday, R. (2011). *A dictionary of media and communication*. Oxford, UK: Oxford University Press. doi:10.1093/acref/9780199568758.001.0001

- Chang, C. Y. (2016). Two decades of research in L2 peer review. *Journal of Writing Research*, 8(1), 81–117. doi:10.17239/jowr-2016.08.01.03
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. London: Sage.
- Chaume, F. (2004). *Cine y traducción*. Madrid, Spain: Cátedra.
- Chaume, F. (2005). Los estándares de calidad y la recepción de la traducción audiovisual. *Puentes*, 6, 5–12.
- Chaume, F. (2007). Quality standards in dubbing: A proposal. *TradTerm*, 13, 71–87. doi:10.11606/issn.2317-9511.tradterm.2007.47466
- Chaume, F. (2012). *Audiovisual translation. Dubbing*. Manchester, UK: St Jerome.
- Cheng, S., & Wu, Q. (2016). Mapping the development of translation competence in learning journals: A dynamic dimension and analytical framework inspired by problem-solving studies. *Chinese Translators Journal*, 1, 51–57.
- Chen, Y. (2010). Trainee translators' reflective learning in small group discussion. In V. Pellatt, K. K. Griffiths, & S. Wu (Eds.), *Teaching and testing interpreting and translating* (pp. 81–95). Bern: Peter Lang.
- Chesterman, A. (1997). Memes of translation. In *Translation strategies*. Philadelphia: John Benjamins. doi:10.1075/btl.22
- Chiaro, D., & Nocella, G. (2004). Interpreters' perception of linguistic and non-linguistic factors affecting quality: A survey through the World Wide Web. *Meta*, 49(2), 278–293. doi:10.7202/009351ar
- Choi, J. Y. (2006). Metacognitive evaluation method in consecutive interpretation for novice learners. *Meta*, 51(2), 273–283. doi:10.7202/013256ar
- Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction*, 20(4), 328–338. doi:10.1016/j.learninstruc.2009.08.006
- Cho, Y. H., & Cho, K. (2011). Peer reviewers learn from giving comments. *An International Journal of the Learning Sciences*, 39(5), 629.
- Clark, J. (2007). *Comments on U.K. guidelines on audio description*. Retrieved from <http://joelclark.org/access/description/ukad.html>
- Clifford, A. (2005). Putting the exam to the test: Psychometric validation and interpreter certification. *Interpreting*, 7(1), 97–131. doi:10.1075/intp.7.1.06cli
- Colina, S. (2008). Translation Quality Evaluation: Empirical Evidence for a Functionalist Approach. *The Translator*, 14(1), 97–134. doi:10.1080/13556509.2008.10799251
- Colina, S. (2009). Further evidence for a functionalist approach to translation quality evaluation. *Target*, 21(1), 97–134.
- Colina, S. (2009). Further evidence for a functionalist approach to translation quality evaluation. *Target: International Journal on Translation Studies*, 21(2), 235–264. doi:10.1075/target.21.2.02col
- Collados Aís, Á. (1998/2002). *La evaluación de la calidad en interpretación simultánea. La importancia de la comunicación no verbal*. Granada, Spain: Comares.
- Collados Aís, A., & García Becerra, O. (2015). Quality criteria. In F. Pöchhacker, N. Grbic, P. Mead, & R. Setton (Eds.), *Encyclopedia of Interpreting Studies* (pp. 368–383). London: Routledge.
- Collados Aís, Á., Iglesias Fernández, E., Pradas Macías, E. M., & Stévaux, E. (2011). *Qualitätsparameter beim Simultandolmetschen: Interdisziplinäre Perspektiven*. Tübingen: Gunter Narr.

Compilation of References

- Collados Aís, Á., Pradas Macías, E. M., Stévaux, E., & García Becerra, O. (Eds.). (2007). *La evaluación de la calidad en interpretación simultánea: parámetros de incidencia*. Granada: Comares.
- Conde. (2009). *Proceso y resultado de la evaluación de traducciones* (Unpublished doctoral dissertation). University of Granada, Granada, Spain. Retrieved October 27, 2017, from <http://www.tdx.cat/handle/10803/16836>
- Corbett, S., LaFrance, M., & Decker, T. E. (Eds.). (2014). *Peer pressure, peer power: Theory and practice in peer review and response for the writing classroom*. Southlake, TX: Fountainhead Press.
- Cornelissen, J. (2014). *Corporate communication: A guide to theory and practice* (4th ed.). London: Sage.
- Corsellis, A., & Vanden Bosch, Y. (2014). Testing knowledge of the legal context. In C. Giambruno (Ed.), *Assessing interpreter quality through testing and certification: The QUALITAS project* (pp. 69–76). Sant Vicent del Raspeig, Spain: Publicaciones de la Universidad de Alicante.
- Darwish, A. (2008). *Optimality in translation*. Victoria: Writescopie Pty Ltd.
- Davidson, N. (1994). Cooperative and collaborative learning. In J. S. Thousand, R. A. Villa, & A. Nevin (Eds.), *Creative and collaborative learning* (pp. 13–30). Baltimore, MD: Paul H. Brookes.
- de Groot, A. M. B. (2000). Complex-skill approach to translation and interpreting. In S. Tirkkonen-Condit & R. Jääskeläinen (Eds.), *Tapping and mapping the process of translation and interpreting* (pp. 53–68). Amsterdam: John Benjamins. doi:10.1075/btl.37.06gro
- de Jongh, E. (2012). *From the classroom to the courtroom: A guide to interpreting in the U.S. justice system*. Amsterdam: John Benjamins. doi:10.1075/ata.xvii
- De los Reyes, J. (2015). *La traducción del cine para niños. Un estudio sobre recepción* (Unpublished doctoral dissertation). Jaume I University, Castellón de la Plana, Spain. Retrieved October 27, 2017, from <http://www.tdx.cat/handle/10803/334991>
- De Yzaguirre, L., García, Y., Matamala Ripoll, A., & Ribas, M. (2005). Tratamiento informático de un corpus de guiones y traducciones de textos audiovisuales: Implicaciones en traductología. In P. Zabalbeascoa, L. Santamaría, & F. Chaume (Eds.), *La traducción audiovisual: Investigación, enseñanza y profesión* (pp. 107–113). Granada, Spain: Comares.
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268. doi:10.1207/S15327965PLI1104_01
- Deci, E. L., & Ryan, R. M. (2002). *Handbook of self-determination research*. Rochester, NY: The University of Rochester Press.
- Déjean Le Féal, K. (1990). Some thoughts on the evaluation of simultaneous interpretation. In D. Bowen & M. Bowen (Eds.), *Interpreting. Yesterday, today, and tomorrow* (pp. 154–160). Binghamton, NY: SUNY. doi:10.1075/ata.iv.27lef
- Delisle, J. (1980). *L'analyse du discours comme méthode de traduction: Théorie et pratique*. Ottawa: University of Ottawa.
- Delisle, J. (1990). *L'enseignement de l'interprétation et de la traduction: De la théorie à la pédagogie*. Ottawa: University of Ottawa.
- Delisle, J. (1993). *La traduction raisonnée: Manuel d'initiation à la traduction professionnelle de l'anglais vers le français*. Ottawa: University of Ottawa.
- Delisle, J. (1998). Définition, rédaction et utilité des objectifs d'apprentissage en enseignement de la traduction. In I. I. Garcia & J. Verdegall (Eds.), *Los estudios de traducción: Un reto didáctico* (pp. 13–44). Castellón: Universitat Jaume I.

- Desjardins, R. (2011). Facebook me!: Initial insights in favour of using social networking as a tool for translator training. *Linguistica Antverpiensia. New Series–Themes in Translation Studies*, 10, 175–196.
- Dewey, J. (1991). *How we think*. New York: Prometheus Books. (Original work published 1910)
- Di Giovanni, E. (2012). Why not ask the audience? An attempt to evaluate viewers' reception of subtitled films at festivals. In F. Dalziel, S. Gesuato, & M. T. Musachio (Eds.), *A lifetime of English studies. Essays in honour of Carol Taylor Torsello* (pp. 471–486). Padua, Italy: Il Poligrafo.
- Di Giovanni, E. (2016). Reception studies in audiovisual translation research. The case of subtitling at film festivals. *Trans-kom*, 9(1), 58–78.
- Díaz-Cintas, J. (2003). *Teoría y práctica de la subtitulación inglés-español*. Barcelona, Spain: Ariel.
- Díaz-Galaz, S., Padilla, P., & Bajo, M. T. (2015). The role of advance preparation in simultaneous interpreting: A comparison of professional interpreters and interpreting students. *Interpreting*, 17(1), 1–25. doi:10.1075/intp.17.1.01dia
- Directorate-General for Translation. European Commission. (2015). *DGT translation quality guidelines*. Retrieved June 13, 2017, from http://ec.europa.eu/translation/maltese/guidelines/documents/dgt_translation_quality_guidelines_en.pdf
- Dodds, R. A., & Smith, S. M. (1999). Fixation. In M. A. Runco & S. R. Pritzker (Eds.), *Encyclopedia of Creativity* (pp. 725–728). San Diego, CA: Academic Press.
- Dragsted, B., & Carl, M. (2013). Towards a classification of translation styles based on eye-racking and keylogging data. *Journal of Writing Research*, 5(1), 133–158. doi:10.1007/978-3-319-20358-4_2
- Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. New York: Free Press.
- Drugan, J. (2013). *Quality in professional translation assessment and improvement*. London: Bloomsbury.
- Drugan, J. (2013). *Quality in professional translation: Assessment and improvement*. London: Bloomsbury.
- Drugan, J. (2014, November). Top-down or bottom-up: What do industry approaches to translation quality mean for effective integration of standards and tools? *Translating and the Computer 36. Proceedings of The International Association for Advancement in Language Technology*.
- Dvorkin, M. (2016). Jobs involving routine tasks aren't growing. *On the Economy, Federal Reserve Bank of St. Louis*. Retrieved June 9, 2017, from <https://www.stlouisfed.org/on-the-economy/2016/january/jobs-involving-routine-tasks-arent-growing>
- Earley, P. C., Northcraft, G. B., Lee, C., & Lituchy, T. R. (1990). Impact of process and outcome feedback on the relation of goal setting to task performance. *Academy of Management Journal*, 33(1), 87–105. doi:10.2307/256353
- Ebel, R. L. (1954). Procedures for the analysis of classroom tests. *Educational and Psychological Measurement*, 1(2), 352–364. doi:10.1177/001316445401400215
- Edmundson, G. (2007). *History of Holland*. Charleston, SC: BiblioLife. Retrieved from <https://www.amazon.com/History-Holland-George-Edmundson/dp/1434625125>
- Ehrensberger-Dow, M., & Massey, G. (2013). Indicators of translation competence: Translators' self-concepts and the translation of titles. *Journal of Writing Research*, 5(1), 103–131. doi:10.17239/jowr-2013.05.01.5

Compilation of References

- Ehrensberger-Dow, M., & Massey, G. (2014). Cognitive ergonomic issues in professional translation. In J. W. Schwieter & A. Ferreira (Eds.), *The development of translation competence: Theories and methodologies from psycholinguistics and cognitive science* (pp. 58–86). Newcastle: Cambridge Scholars.
- Ehrensberger-Dow, M., & Massey, G. (2017). Socio-technical issues in professional translation practice. *Translation Spaces*, 6(1), 104–121. doi:10.1075/ts.6.1.06ehr
- Ehrensberger-Dow, M., & O'Brien, S. (2015). Ergonomics of the translation workplace: Potential for cognitive friction. *Translation Spaces*, 4(1), 98–118. doi:10.1075/ts.4.1.05ehr
- Elbow, P. (1973). *Writing without teachers*. Oxford, UK: Oxford University Press.
- Elbow, P. (1981). *Writing with power*. Oxford, UK: Oxford University Press.
- Elkington, S. (2016). *HEA transforming assessment in higher education summit 2016: Final report*. Retrieved July 20, 2017, from https://www.heacademy.ac.uk/system/files/downloads/hea_assessment_summit_2016_0.pdf
- EMT Expert Group. (2009). *Competences for professional translators, experts in multilingual and multimedia communication*. Brussels: European Commission, Directorate-General for Translation. Retrieved October 30, 2016, from http://ec.europa.eu/dgs/translation/programmes/emt/key_documents/emt_competences_translators_en.pdf
- EMT Expert Group. (2009). *Competences for professional translators, experts in multilingual and multimedia communication*. Retrieved January, 8, 2018, from https://ec.europa.eu/info/sites/info/files/emt_competences_translators_en.pdf
- EMT Expert Group. (2009). *Competences for professional translators, experts in multilingual and multimedia communication*. Retrieved July 11, 2017, from https://ec.europa.eu/info/sites/info/files/emt_competences_translators_en.pdf
- EMT Expert Group. (2009). *Competences for professional translators, experts in multilingual and multimedia communication*. Retrieved June 13, 2017, from https://ec.europa.eu/info/education/european-masters-translation-emt/european-masters-translation-emt-explained_en
- EMT Expert Group. (2013). *The EMT translator trainer profile competences of the trainer in translation*. Retrieved June 13, 2017, from <http://docplayer.net/14445272-The-emt-translator-trainer-profile-competences-of-the-trainer-in-translation.html>
- EN 15038. (2006). *Translation services — service requirements*. Brussels: European Committee for Standardization.
- Englund Dimitrova, B., & Tiselius, E. (2009). Exploring retrospection as a research method for studying the translation process and the interpreting process. In I. M. Mees, F. Alves & S. Göpferich (Eds.), *Methodology, technology and innovation in translation process research: A tribute to Arnt Lykke Jakobsen* (pp. 109-134). Fredriksberg: Samfundslitteratur.
- Englund Dimitrova, B. (2005). *Expertise and explicitation in the translation process*. Amsterdam: John Benjamins. doi:10.1075/btl.64
- Ericsson, K. A. (1996). The Acquisition of Expert Performance: An Introduction to Some of the Issues. In K. A. Ericsson (Ed.), *The Acquisition of Expert Performance in the Arts and Sciences, Sports and Games* (pp. 1–50). Mahwah, NJ: Lawrence Erlbaum Associates.
- Eskelinen, J., & Pakkala-Weckström, M. (2016). Assessing translation students' learning of professional competences. *Translation Spaces*, 5(2), 314–331. doi:10.1075/ts.5.2.08esk
- European Commission. Directorate-General for Translation. (2012). *The status of the translation profession in the European Union. Final report*. Luxembourg: Publications Office of the European Union.

- European Master's in Translation (EMT). (2009). *Competences for professional translators, experts in multilingual and multimedia communication*. Retrieved January 5, 2018, from https://ec.europa.eu/info/sites/info/files/emt_competences_translators_en.pdf
- Evans, A. W., McKenna, C., & Oliver, M. (2005). Trainees' perspectives on the assessment and self-assessment of surgical skills. *Assessment & Evaluation in Higher Education*, 30(2), 163–174. doi:10.1080/0260293042000264253
- Even-Zohar, I. (1990). The position of translated literature within the literary polysystem. *Poetics Today*, 11(1), 45–51. doi:10.2307/1772668
- Eyckmans, J., Anckaert, P., & Segers, W. (2009). The perks of norm-referenced translation evaluation. In C. V. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting Studies: A call for dialogue between research and practice* (pp. 73–93). Amsterdam: John Benjamins. doi:10.1075/ata.xiv.06eyc
- Eyckmans, J., Anckaert, P., & Segers, W. (2016). Translation and interpretation skills. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 219–235). Boston: Walter De Gruyter. doi:10.1515/9781614513827-016
- Faculty Innovation Center, University of Texas at Austin. (n.d.). *Peer assessment*. Retrieved July 6, 2017, from <https://facultyinnovate.utexas.edu/teaching/check-learning/feedback/peer-assessment>
- Faigley, L., & Witte, S. (1981). Analyzing revision. *College Composition and Communication*, 32(4), 400–414. doi:10.2307/356602
- Falchikov, N. (2005). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. London: Routledge-Falmer.
- Feedback. (2017). In *Merriam-Webster.com*. Retrieved October 14, 2017, from <http://www.merriam-webster.com/dictionary/feedback>
- Feedforward. (2017). In *Oxforddictionaries.com*. Retrieved October 14, 2017, from <https://en.oxforddictionaries.com/definition/feedforward>
- Feuerle, L. M. (2013). Testing interpreters: Developing, administering, and scoring court interpreter certification exams. *The International Journal of Translation and Interpreting Research*, 5(1), 297–323. doi:10.12807/ti.105201.2013.a04
- Fiederer, R., & O'Brien, S. (2009). Quality and machine translation: A realistic objective? *Journal of Specialised Translation*, 11, 52–74.
- Fields, P., Hague, D., Koby, G., Lommel, A., & Melby, A. K. (2014). What is quality? A management discipline and the translation industry get acquainted. *Revista Tradumàtica: tecnologies de la traducció*, 12, 404–412.
- Fitzgerald, J. (1992). *Knowledge in writing: Illustration from revision studies*. New York: Springer. doi:10.1007/978-1-4612-2796-0
- Flanagan, M., & Heine, C. (2015). Peer-feedback as a translation training tool in Web-base communication. *Hermes*, 54, 115–136.
- Flanagan, M., & Heine, C. (2015). Peer-feedback as a translation training tool in web-based communication. *Hermes*, 54, 115–136.
- Flanagan, M., & Heine, C. (2017). Peer-feedback as a translation training tool in web-based communication. *Journal of Language and Communication in Business*, 54, 115–136.

Compilation of References

- Flower, L. S. (1979). Writer-based prose: A cognitive basis for problems in writing. *College English*, 41(1), 19–37. doi:10.2307/376357
- Fowler, Y. (2007). Formative assessment: Using peer and self-assessment in interpreter training. In *The critical link 4: Professionalisation of interpreting in the community* (pp. 253–262). Amsterdam: John Benjamins.
- Fox, J. (2003). Effect Displays in {R} for Generalised Linear Models. *Journal of Statistical Software*, 8(15), 1–27. doi:10.18637/jss.v008.i15
- Fox, O. (2000). The use of translation diaries in a process-oriented translation teaching methodology. In C. Schäffner & B. Adab (Eds.), *Developing translation competence* (pp. 115–130). Amsterdam: John Benjamins. doi:10.1075/btl.38.12fox
- Fryer, L., & Freeman, J. (2012a). Cinematic language and the description of film: Keeping AD users in the frame. *Perspectives: Studies in Translatology*. doi:10.1080/0907676X.2012.693108
- Fryer, L., & Freeman, J. (2013). Visual impairment and presence: Measuring the effect of audio description. In H. Witchel (Ed.), *Inputs-Outputs Conference. Proceedings: Interdisciplinary approaches to causality in engagement, immersion and presence in performance and human-computer interaction* (pp. 1–5). Brighton, UK: Academic Press.
- Fryer, L., & Freeman, J. (2014). Can you feel what I'm saying? The impact of verbal information on emotion elicitation and presence in people with a visual impairment. In A. Felnhofner & O. D. Kothgassner (Eds.), *Challenging presence: Proceedings of the international society for presence research, 15th International Conference on Presence* (pp. 99 – 107). Vienna: Facultas.
- Fryer, L. (2016). *An introduction to audio description: A practical guide*. London: Routledge.
- Fryer, L. (in press). Perception and psycholinguistics in AVT. In L. Perez-Gonzalez (Ed.), *The Routledge handbook of audio visual translation*. London: Routledge.
- Fryer, L. E. (2009). *Talking dance: The audio describer's guide to dance in theatre*. London: Audio Description Association.
- Fryer, L., & Freeman, J. (2012b). Presence in those with and without sight: Implications for virtual reality and audio description. *Journal of Cyber Therapy and Rehabilitation*, 5, 15–23.
- Fuentes Luque, A. (2001). *La recepción del humor audiovisual traducido: Estudio comparativo de fragmentos de las versiones doblada y subtitulada al español de la película «Duck Soup» de los Hermanos Marx* (Unpublished doctoral dissertation). University of Granada, Granada, Spain. Retrieved October 27, 2017, from <http://www.tdx.cat/handle/10803/145573>
- Fuentes Luque, A. (2005). La recepción de la traducción audiovisual: Humor y cultura. In P. Zabalbeascoa, L. Santamaría, & F. Chaume (Eds.), *La traducción audiovisual: Investigación, enseñanza y profesión* (pp. 139–152). Granada, Spain: Comares.
- Gaddis Rose, M. (1979). Translation as problem-solving. *Translation Review*, 3(1), 20–21. doi:10.1080/07374836.1979.10523556
- Galán-Mañas, A. (2016). Learning portfolio in translator training: The tool of choice for competence development and assessment. *The Interpreter and Translator Trainer*, 10(2), 161–182. doi:10.1080/1750399X.2015.1103108
- Galán-Mañas, A., & Hurtado Albir, A. (2015). Competence assessment procedures in translator training. *The Interpreter and Translator Trainer*, 9(1), 63–82. doi:10.1080/1750399X.2015.1010358
- Gambier, Y. (2012). Teaching translation / Training translators. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of Translation Studies* (Vol. 3, pp. 163–171). Amsterdam: John Benjamins. doi:10.1075/hts.3.tea1

- Garant, M. (2009). A case for holistic assessment. *AFinLA-e Soveltavan Kielittiteen Tutkimuksia*, 1, 5-17. Retrieved from <http://hdl.handle.net/10138/154545>
- Garant, M. (2009). A case for holistic translation assessment. *AFinLA-e: Soveltavan kielittiteen tutkimuksia*, 1, 5–17.
- Garant, J., & Garant, M. (2001). Case study: Variation of assessment practices in translator education. In P. Kukkonen & R. Hartama-Heinonen (Eds.), *Mission, vision, strategies, and values: A celebration of translator training and translator studies in Kouvola* (pp. 45–58). Helsinki: Helsinki University Press.
- Garber, N., & Mauffette-Leenders, L. A. (1997). Obtaining feedback from non-English speakers. In S. E. Carr, R. Roberts, A. Dufour, & D. Steyn (Eds.), *The critical link: Interpreters in the community* (pp. 131–146). Amsterdam: John Benjamins. doi:10.1075/btl.19.15gar
- García Álvarez, A. M. (2007). Evaluating students' translation process in specialised translation: Translation commentary. *The Journal of Specialised Translation*, 7, 139–163.
- Garvin, D. G. (1984). What does product quality really mean? *Sloan Management Review*, 26(1), 25–43.
- Garvin, D. G. (1987). Competing on the eight dimensions of quality. *Harvard Business Review*, 65(6), 101–109.
- Giambruno (Miguélez), C. (1997). *Language mediation in the judicial system: The role of the court interpreter* (Unpublished doctoral dissertation). Universitat d'Alacant, Spain. Retrieved from file:///Users/dqf115/Downloads/Giambruno-Cynthia.pdf
- Giambruno, C. (2016). Quality assessment and political will. *Language and Law/Linguagem E Direito*, 3(2), 116-134.
- Giambruno, C. (2014). *Assessing legal interpreter quality through testing and certification: The QUALITAS project*. San Vicent del Raspeig, Spain: Publicaciones de la Universidad de Alicante.
- Giambruno, C. (2014). The current state of affairs in the UE: Member state profiles. In C. Giambruno (Ed.), *Assessing interpreter quality through testing and certification: The QUALITAS project* (pp. 149–190). Sant Vicent del Raspeig, Spain: Publicaciones de la Universidad de Alicante.
- Gile, D. (1991). A Communication-Oriented Analysis of Quality. In M. L. Larson (Ed.), *Translation: Theory and practice* (pp. 188-200). Binghamton, NY: SUNY.
- Gile, D. (1994). The Process-Oriented Approach in Translation Training. In C. Dollerup & A. Lindegaard (Eds.), *Teaching Translation and Interpreting 2: Insights, Aims, Visions* (pp. 107-112). Amsterdam: John Benjamins. doi:10.1075/btl.5.17gil
- Gile, D. (1991). A communicative-oriented analysis of quality in nonliterary translation and interpretation. In M. L. Larson (Ed.), *Translation: Theory and practice, Tension and interdependence* (pp. 188–200). Binghamton, NY: SUNY. doi:10.1075/ata.v.19gil
- Gile, D. (1999). Variability in the perception of fidelity in simultaneous interpretation. *Hermes*, 22, 51–79.
- Gile, D. (2001). L'évaluation de la qualité de l'interprétation en cours de formation. *Meta: Journal des traducteurs'/ Meta: Translators' Journal*, 46(2), 379–393.
- Gile, D. (2004). Integrated problem and decision reporting as a translator training tool. *The Journal of Specialised Translation*, 2, 2–20.
- Gile, D. (2009). *Basic concepts and models for interpreter and translator training* (Revised ed.). Amsterdam: John Benjamins. doi:10.1075/btl.8

Compilation of References

- Gnecchi, M., Maylath, B., Scarpa, F., Mousten, B., & Vandepitte, S. (2011). Field convergence: Merging roles of technical writers and technical translators. *IEEE Transactions on Professional Communication*, *54*, 168–184. doi:.212175010.1109/TPC.2011
- Goldin, G. A. (1997). Observing mathematical problem solving through task-based interviews. *Qualitative Research Methods in Mathematics Education*, *9*, 40–62.
- González Davies, M. (2004). *Multiple voices in the translation classroom: Activities, tasks and projects*. Amsterdam: John Benjamins. doi:10.1075/btl.54
- González-Hibner, M., Palma, J., & Korp-Edwards, B. (2017). *From the NAJIT board: The future of the NJITCE*. Retrieved from <https://najit.org/from-the-najit-board-the-future-of-the-njitce/>
- Göpferich, S. (2008). *Translationsprozessforschung. Stand - Methoden - Perspektiven*. Tübingen: Narr.
- Göpferich, S. (2009). Towards a model of translation competence and its acquisition: The longitudinal study TransComp. In S. Göpferich, A. L. Jakobsen, & I. M. Mees (Eds.), *Behind the mind: Methods, models and results in translation process research* (pp. 11–37). Copenhagen: Samfundslitteratur Press.
- Göpferich, S. (2012). Tracing strategic behaviour in translation processes: Translation novices, 4th-semester students and professional translators compared. In S. Hubscher-Davidson & M. Borodo (Eds.), *Global trends in translator and interpreter training* (pp. 240–266). London: Continuum.
- Göpferich, S. (2013). Translation competence: Explaining development and stagnation from a dynamic systems perspective. *Target*, *25*(1), 61–76. doi:10.1075/target.25.1.06goe
- Göpferich, S., & Jääskeläinen, R. (2009). Process research into the development of translation competence: Where are we, and where do we need to go? *Across Languages and Cultures*, *10*(2), 169–191. doi:10.1556/Acr.10.2009.2.1
- Görög, A. (2014). Quantification and comparative evaluation of quality: The TAUS dynamic quality framework. *Revista Tradumática*, *12*, 443–454. doi:10.5565/rev/tradumatica.66
- Gouadec, D. (2007). *Translation as a profession*. Amsterdam: John Benjamins. doi:10.1075/btl.73
- Gouadec, D. (2010). Quality in translation. In Y. Gambier & L. Van Doorslaer (Eds.), *Handbook of translation studies* (Vol. 1, pp. 270–275). Amsterdam: John Benjamins. doi:10.1075/hts.1.qual
- Gouadec, D. (2013). Quality in translation. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of Translation Studies* (pp. 271–275). Amsterdam: John Benjamins.
- Grace, K., Salvatier, J., Dafoe, A., Zhang, B., & Evans, O. (2017). *When will AI exceed human performance? Evidence from AI experts*. Retrieved June 9, 2017, from <https://arxiv.org/pdf/1705.08807.pdf>
- Graham, J. F. (1985). *Difference in translation*. London: Cornell University Press.
- Greco, G. M. (2017). *Understanding media accessibility quality (UMAQ)*. Paper presented at Media Accessibility Platform International Conference (MAPIC), Vigo.
- Greene, B. (2011). *The elegant universe: Superstrings, hidden dimensions, and the quest for the ultimate theory*. New York: Vintage Books.
- Gries, S. T. (2015). Some current quantitative problems in corpus linguistics and a sketch of some solutions. *Language and Linguistics (Taipei)*, *16*(1), 93–117. doi:10.1177/1606822X14556606

- Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition and Emotion*, 9(1), 87–108. doi:10.1080/02699939508408966
- Guerberof Arenas, A. (2009). Productivity and quality in the post-editing of outputs from translation memories and machine translation. *Localisation Focus*, 7(1), 11–21.
- Guerberof Arenas, A. (2013). What do professional translators think about post-editing? *Journal of Specialised Translation*, 19, 75–95.
- Guerberof Arenas, A. (2014a). The role of professional experience in post-editing from a quality and productivity perspective. In S. O'Brien, L. Winther Balling, M. Carl, M. Simard, & L. Specia (Eds.), *Post-editing of machine translation: Processes and applications* (pp. 51–76). Newcastle: Cambridge Scholars.
- Guerberof Arenas, A. (2014b). Correlations between productivity and quality when post-editing in a professional context. *Machine Translation*, 28(3-4), 165–186. doi:10.1007/10590-014-9155-y
- Habermas, J. (1971). *Knowledge and human interests*. Boston: Beacon Press.
- Hagemann, S. (2016). (Non-)professional, authentic projects? Why terminology matters. In D. Kiraly, L. R uth, C. Canfora & A. Cnyrim (Eds.), *Towards authentic experiential learning in translator education* (pp. 33-51). G ttingen: V&R unipress/Mainz University Press.
- Han, C. (2017). Using analytic rating scales to assess English-Chinese bi-directional interpretation: A longitudinal Rasch analysis of scale utility and rater behavior. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 16.
- Han, C. (2018). *A longitudinal quantitative investigation into the validity of self and peer assessment applied to English-Chinese bi-directional interpretation in an undergraduate interpreting course*. Manuscript submitted for publication.
- Han, C. (2015). Investigating rater severity/leniency in interpreter performance testing: A multifaceted Rasch measurement approach. *Interpreting*, 17(2), 255–283. doi:10.1075/intp.17.2.05han
- Han, C. (2016). Investigating score dependability in English/Chinese interpreter certification performance testing: A generalizability theory approach. *Language Assessment Quarterly*, 13(3), 186–201. doi:10.1080/15434303.2016.1211132
- Han, C., & Riazi, M. (2017). Investigating the effects of speech rate and accent on simultaneous interpretation: A mixed-methods approach. *Across Languages and Cultures*, 18(2), 237–259. doi:10.1556/084.2017.18.2.4
- Han, C., & Riazi, M. (2018). The accuracy of student self-assessments of English-Chinese bidirectional interpretation: A longitudinal quantitative study. *Assessment & Evaluation in Higher Education*, 1–13. doi:10.1080/02602938.2017.1353062
- Han, C., & Slatyer, H. (2016). Test validation in interpreter certification performance testing: An argument-based approach. *Interpreting*, 18(2), 231–258. doi:10.1075/intp.18.2.04han
- Hansen, G. (2006). Retrospection Methods in Translator Training and Translation Research. *Journal of Specialized Translation*, 5, 2–40.
- Hansen, G. (2010). Integrative description of translation processes. In G. M. Shreve & E. Angelone (Eds.), *Translation and cognition* (pp. 189–211). Amsterdam: John Benjamins. doi:10.1075/ata.xv.11han
- Hartley, A., Mason, I., Peng, G., & Perez, I. (2003). *Peer- and self-assessment in conference interpreter training*. Retrieved from <https://www.llas.ac.uk/resourcedownloads/1454/hartley.rtf>
- Hattie, J., Gan, M., & Brooks, C. (2011). Instruction based on feedback. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 290–324). New York: Routledge.

Compilation of References

- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. doi:10.3102/003465430298487
- Hatton, N., & Smith, D. (1995). Reflection in teacher education – towards definition and implementation. *Teaching and Teacher Education*, 11(1), 33–49. doi:10.1016/0742-051X(94)00012-U
- Hayes, J. R. (1989). *The complete problem solver* (2nd ed.). Lawrence Erlbaum Associates, Inc.
- Heimeri-Moggan, K. (2016). Fit-to-practise versus exam-ready legal public service interpreter training: Training for real life or for the exam? In T. M., G. Webb & M. Rabadán-Gómez (Eds.), *Challenges and opportunities in public service interpreting*. Academic Press. doi:10.1057/978-1-137-45000-5_5
- Henry, J., & Ledbetter, L. (2011). Teaching intellectual teamwork in WAC courses through peer review. *Currents in Teaching and Learning*, 3(2), 4–21.
- Herbert, J. (1952). *The Interpreter's Handbook*. Geneva: Georg.
- Hermans, T. (1991). Translation norms and correct translations. In K. van Leuven-Zwart & T. Naaijken (Eds.), *Translation studies: The state of the art. Proceedings of the first James S. Holmes Symposium on Translation Studies* (pp. 155–169). Amsterdam: Rodopi.
- Hertog, E. (2014). Testing knowledge of codes of ethics. In C. Giambruno (Ed.), *Assessing legal interpreter quality through testing and certification: The QUALITAS project* (pp. 77–92). Sant Vicent del Raspeig, Spain: Publicaciones de la Universidad de Alicante.
- Higher Education Academy (HEA). (2012). *A marked improvement: Transforming assessment in Higher Education*. Retrieved June 13, 2017, from https://www.heacademy.ac.uk/system/files/a_marked_improvement.pdf
- Higher Education Funding Council for England (HEFCE). (2016). *National Student Survey Results 2016*. Retrieved June 13, 2017, from <http://www.hefce.ac.uk/lt/nss/results/2016/>
- Hlavac, J. (2013). A cross-national overview of translator and interpreter certification procedures. *The International Journal of Translation and Interpreting Research*, 5(1), 32–65. doi:10.12807/ti.105201.2013.a02
- Hlavac, J., & Orlando, M. (2015). 'Capturing' and 'prescribing' desirable attributes of community interpreting: Design and implementation of testing systems in four anglophone countries. In C. Zwischenberger & M. Behr (Eds.), *Interpreting quality: A look around and ahead* (pp. 297–324). Berlin: Frank & Timme.
- Hofmann, S. (2012). *Prozessgestütztes Übersetzen. Vom funktionsorientierten Übersetzungsprozess zum GP-Modell für die Dienstleistung Übersetzen*. Lichtenberg: Harland Media.
- Holec, H. (1981). *Autonomy and foreign language learning*. Oxford, UK: Pergamon Press. (first published in 1979 for the Council of Europe)
- Holmes, J. S. (1988). *Translated! Papers on literary translation and translation studies*. Amsterdam: Rodopi.
- Holroyd, C. (2000). Are assessors professionals? *Active Learning in Higher Education*, 1(1), 28–44. doi:10.1177/1469787400001001003
- Holtzhausen, D., & Zerfass, A. (2015). *The Routledge handbook of strategic communication*. New York: Routledge.
- Hönig, H. G., & Kussmaul, P. (1982). *Strategie der Übersetzung. Ein Lehr- und Arbeitsbuch*. Tübingen: Narr.
- Hönig, H., & Kussmaul, P. (1982). *Strategie der Übersetzung: Ein Lehr- und Arbeitsbuch*. Tübingen: Narr.

- Hounsell, D. (1997/2005). Contrasting conceptions of essay-writing. In F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The Experience of learning: Implications for teaching and studying in higher education* (pp. 106–125). Edinburgh, UK: University of Edinburgh, Centre for Teaching, Learning and Assessment.
- House, J. (2015). *Translation quality assessment. Past and present*. Oxon, UK: Routledge. Retrieved June 25, 2017, from <http://www.translationindustry.ir/Uploads/Pdf/Translation%20Quality%20Assessment.pdf>
- House, J. (1997). *Translation quality assessment: A model revisited*. Tübingen: Gunter Narr Verlag.
- House, J. (2015). *Translation quality assessment: Past and present*. London: Routledge.
- Hoyle, D. (2001). *ISO 9000 quality systems handbook* (4th ed.). Oxford, UK: Butterworth Heinemann. Retrieved June 11, 2017, from <http://www.pqm-online.com/assets/files/lib/books/holye2.pdf>
- Huahui, Z. (2010). Investigating learner's use and understanding of peer and teacher feedback on writing: A comparative study in a Chinese English writing classroom. *Assessing Writing*, 13(1), 3–17.
- Huck-Sandhu, S. (2015). Internationale PR-Arbeit. In G. Bentele, R. Fröhlich, & P. Szyszka (Eds.), *Handbuch Public Relations. Wissenschaftliche Grundlagen und berufliches Handeln* (pp. 757–772). Wiesbaden: Springer VS.
- Huck-Sandhu, S. (2016). Internationale Unternehmenskommunikation. In C. Mast (Ed.), *Unternehmenskommunikation. Ein Leitfaden* (pp. 423–440). Konstanz: UVK.
- Huertas Barros, E. (2011). Collaborative learning in the translation classroom: Preliminary survey results. *The Journal of Specialised Translation*, 16. Retrieved from http://www.jostrans.org/issue16/art_barros.php
- Huertas Barros, E. (2013). *La competencia interpersonal en la formación de traductores en España: Un estudio empírico-descriptivo del trabajo colaborativo durante la primera etapa de formación en Traducción e Interpretación* (Doctoral dissertation). Retrieved December 28, 2017, from <https://hera.ugr.es/tesisugr/22074508.pdf>
- Huertas Barros, E., & Vine, J. (2015). Assessing the situation: Closing the gap in translation assessment. *The Linguist*, 54(4), 22–24. Retrieved December 28, 2017, from <http://thelinguist.uberflip.com/i/552383-the-linguist-54-4/21>
- Huertas Barros, E., & Vine, J. (2018). Current trends on MA translation courses in the UK: Changing assessment practices on core translation modules. *The Interpreter and Translator Trainer*, 12(1).
- Huertas Barros, E., & Vine, J. (2018). Current trends on MA translation courses in the UK: Changing assessment practices on core translation modules. *The Interpreter and Translator Trainer*, 12(1). Retrieved from <http://www.tandfonline.com/doi/full/10.1080/1750399X.2017.1400365>
- Huertas Barros, E., & Vine, J. (2018a). Current trends on MA translation courses in the UK: Changing assessment practices on core translation modules. *The Interpreter and Translator Trainer*, 12(1). Retrieved from <http://www.tandfonline.com/doi/full/10.1080/1750399X.2017.1400365>
- Huertas Barros, E., & Vine, J. (2018b). Constructing standards in communities: Tutors' and students' perceptions of assessment practices on an MA translation course. In E. Huertas Barros, S. Vandepitte & E. Iglesias Fernández (Eds.), *Quality assurance and assessment practices in translation and interpreting: Advances in linguistics and communication studies (ALCS)*. Hershey, PA: IGI Global.
- Huertas Barros, E. (2011). Collaborative learning in the translation classroom: Preliminary survey results. *The Journal of Specialised Translation*, 16, 42–60.

Compilation of References

- Huertas Barros, E., & Buendía Castro, M. (2018). Analysing phraseological units in legal translation: Evaluation of translation errors for the English-Spanish language pair. In S. Gozdz-Roszkowski & G. Pontrandolfo (Eds.), *Phraseology in legal and institutional settings: A corpus-based interdisciplinary perspective* (pp. 41–60). London: Routledge.
- Huertas Barros, E., & Vine, J. (2016). Translator trainers' perceptions of assessment: An empirical study. In M. Thelen, G.-W. van Egdom, D. Verbeeck, Ł. Bogucki, & B. Lewandowska-Tomaszczyk (Eds.), *Translation and meaning: New Series, 41* (pp. 29–39). Frankfurt am Main: Peter Lang.
- Humbley, J., Maylath, B., Mousten, B., Vandepitte, S., & Veisblat, L. (2005). Learning localization through trans-Atlantic collaboration. In G. F. Hayhoe (Ed.), *Professional communication conference proceedings, 2005. IPCC 2005. IEEE International Proceedings* (pp. 578-595). New York: IEEE. 10.1109/IPCC.2005.1494227
- Hurtado Albir, A. (1999/2003). *Enseñar a traducir: Metodología en la formación de traductores e intérpretes*. Madrid: Edelsa.
- Hurtado Albir, A. (2001/2004). *Traducción y traductología: Introducción a la traductología*. Madrid: Cátedra.
- Hurtado Albir, A. (2015a). *Aprender a traducir del francés al español (guía didáctica): Competencias y tareas para la iniciación a la traducción*. Castelló de la Plana: Universidad Jaume I. Servicio De Comunicación y Publicaciones.
- Hurtado Albir, A. (2015b). The acquisition of translation competence: Competences, tasks, and assessment in translator training. *Meta: Journal des Traducteurs / Meta: Translators' Journal*, 60(2), 256–280.
- Hurtado Albir, A. (2017). Translation and translation competence. In A. Hurtado Albir (Ed.), *Researching translation competence by PACTE Group* (pp. 3–34). Amsterdam: John Benjamins. doi:10.1075/btl.127.01hur
- Hurtado Albir, A. (Ed.). (2017). *Researching translation competence by PACTE Group*. Amsterdam: John Benjamins. doi:10.1075/btl.127
- Hurtado Albir, A., & Taylor, P. (2015). The Acquisition of Translation Competence. Competences, Tasks and Assessment in Translator Training. *Meta*, 60(2), 256–280. doi:10.7202/1032857ar
- Hurtado, A. (2001). *Traducción y traductología. Introducción a la traductología* (6th ed.). Madrid, Spain: Cátedra.
- Hurtado-Albir, A. (2007). Competence-based curriculum design for training translators. *The Interpreter and Translator Trainer*, 1(2), 163–195. doi:10.1080/1750399X.2007.10798757
- Iaroslavski, M. (2011). Becoming an interpreter: Assessment and self-assessment in untutored practice sessions. *Cognition, Brain, Behavior: An Interdisciplinary Journal*, 15(2), 229–249.
- Ibrahim-González, N., & Noordin, N. (2012). Learner autonomy via self-assessment in consecutive interpreting for novice learners in a non-interpreting environment. *Sino-US English Teaching*, 9(4), 1065–1073.
- Iglesias Fernández, E. (2011). Under examination. Do all interpreting examiners use the same criteria? *The Linguist*, 50-2. Retrieved from https://www.researchgate.net/publication/266317103_Under_Examination_Do_All_Interpreting_Examiners_Use_The_Same_Criteria
- Iglesias Fernández, E. (2013). Unpacking delivery criteria in interpreting quality assessment. In D. Tsagari & R. van Deemter (Eds.), *Assessment issues in language, translation and interpreting* (pp. 51-56). Frankfurt am Main: Peter Lang. Retrieved from <http://interpspracticegroup.com/about-us>
- Iglesias Fernández, E. (2006). La indefinición del parámetro “agradabilidad de la voz” y los estudios de calidad de la interpretación simultánea. In M. J. Varela (Ed.), *La evaluación en los estudios de traducción e interpretación* (pp. 225–239). Seville, Spain: Bienza.

- Iglesias Fernández, E. (2013). Unpacking delivery criteria in interpreting quality assessment. In D. Tsagari & R. van Deemter (Eds.), *Assessment issues in language translation and interpreting* (pp. 51–66). Frankfurt: Peter Lang.
- Iglesias Fernández, E., Martínez Martínez, S., & Chica Núñez, A. J. C. (2015). Cross-fertilization between reception studies in audio description and interpreting quality assessment: The role of the describer's voice. In R. Baños Piñero & J. Díaz Cintas (Eds.), *Audiovisual translation in a global context. Mapping an ever-changing landscape* (pp. 72–95). Palgrave Macmillan, UK.
- International, S. A. E. (2005). *Translation quality metrics (J2450)*. Warrendale, PA: SAE.
- InterpretimeBank. (n.d.). Retrieved from <http://interpretimebank.net>
- Ipsen, A. H., & Dam, H. V. (2016). Translation revision: Correlating revision procedure and error detection. *Hermes: Journal of Language and Communication in Business*, 55, 143–156.
- ISO 17100 (2015). *Translation services — Requirements for translation services*. Geneva: ISO.
- ISO 17100 (2015). *Translation services: Requirements for translation services (ISO 17100:2015)*; English. Berlin: Beuth.
- ISO 17100. (2015). *Translation services — Requirements for translation services*. Geneva: ISO.
- ISO 17100. (2015). *Translation services: Requirements for translation services*. Geneva: ISO. Retrieved from <https://www.iso.org/standard/59149.html>
- ISO 9001 (2015). *Quality management systems: Requirements. Chapter 8 (ISO 9001:2015)*; German and English versions. Berlin: Beuth.
- ISO. (2005). *Quality management systems: Fundamentals and vocabulary, ISO 9000:2005(E)*. Geneva: ISO.
- ISO. (2015). *ISO 17100 Translation services: Requirements for translation services*. Geneva: ISO.
- ISO. (2015). *Translation Services – Requirements for translation services. ISO 17100:2015*. Geneva: ISO. Retrieved from <https://www.iso.org/standard/59149.html>
- ISO. (2015). *Translation services: Requirements for translation services, ISO 17100:2015*. Geneva: ISO.
- ITC. (2000). *ITC guidance on standards for audio description*. Retrieved from http://www.ofcom.org.uk/static/archive/itc/itc_publications/codes_guidance/audio_description/index.asp. html
- J.D. Edwards QA Form. (n.d.). Retrieved July 4, 2017, from <https://www.slideshare.net/patigalin/translation-quality-measurement2-6863666>
- Jääskeläinen, R. (2002). Think-aloud protocol studies. *Target*, 14(1), 107–136. doi:10.1075/target.14.1.05jaa
- Jacobson, H. E. (2009). Moving beyond words in assessing mediated interaction. Measuring interactional competence in healthcare settings. In C. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies* (pp. 49–70). Amsterdam: John Benjamins. doi:10.1075/ata.xiv.04jac
- Jakobsen, A. L. (2002). Translation drafting by professional translators and by translation students. In G. Hansen (Ed.), *Empirical Translation Studies: process and product* (pp. 191–204). Copenhagen: Samfundslitteratur.
- Jakobsen, A. L. (2003). Effects of think aloud on translation speed, revision and segmentation. In F. Alves (Ed.), *Triangulating Translation: Perspectives in Process Oriented Research*, 2 (pp. 69–95). Amsterdam: John Benjamins; doi:10.1075/btl.45.08jak

Compilation of References

- Jekat, S. J., Prontera, D., & Bale, R. (2015). On the perception of audio description: Developing a model to compare films and their audio described versions. *Trans-kom*, 8(2), 446–464.
- Johnson, J. (2003). Learning through portfolios in the translation classroom. In J. Baer & G. S. Koby (Eds.), *Beyond the ivory tower: Rethinking translation pedagogy* (pp. 97–116). Amsterdam: John Benjamins. doi:10.1075/ata.xii.08joh
- Jolley, J. (2017, July/August). Can competency-based education enhance translator training? *The ATA Chronicle*, 46, 15-17. Retrieved from <http://www.atanet.org/chronicle-online/featured/can-competency-based-education-enhance-translator-training/#sthash.UyFrvLON.dpbs>
- Jones, R. (1998). *Conference interpreting explained*. Manchester, UK: St. Jerome Publishing.
- Kalina, S. (2005). Quality assurance for interpreting processes. *Meta: Journal des traducteurs/ Meta: Translators' Journal*, 50(2), 768–784.
- Katan, D. (2011). Occupation or profession: A survey of the translators' world. In R. Sela-Sheffy & M. Shlesinger (Eds.), *Profession, identity and status: Translators and interpreters as an occupational group* (pp. 65–88). Amsterdam: Benjamins. doi:10.1075/bct.32.06kat
- Katan, D. (2013). Cultural mediation. In Y. Gambier & D. Van Doorslaer (Eds.), *Handbook of translation studies*, 4 (pp. 84–91). Amsterdam: Benjamins. doi:10.1075/hts.4.int5
- Katan, D. (2016). Translation at the cross-roads: Time for the transcultural turn? *Perspectives*, 24(3), 365–381. doi:10.1080/0907676X.2015.1016049
- Kearney, J. (2013). Improving engagement: The use of 'authentic self and peer-assessment for learning' to enhance the student learning experience. *Assessment & Evaluation in Higher Education*, 38(7), 875–891. doi:10.1080/02602938.2012.751963
- Kelly, D. (2002). Un modelo de competencia traductora: Bases para el diseño curricular. *Puentes*, 1, 9–20.
- Kelly, D. (2005). *A handbook for translator trainers*. Manchester, UK: St. Jerome.
- Kelly, D. (2005). *A Handbook for translator trainers*. Manchester: St. Jerome.
- Kelly, D. (2005). *A handbook for translator trainers: A guide to reflective practice*. Manchester: St Jerome Publishing.
- Kelly, D. (2005). *A Handbook for translator trainers: Translation practices explained*. Manchester, UK: St. Jerome.
- Kelly, D. (2005). *A handbook for translator trainers: Translation practices explained*. Manchester: St. Jerome.
- Kelly, D. (2007). Translator competence contextualized. Translator training in the framework of higher education reform: In search of alignment in curricular design. In D. Kenny & K. Ryou (Eds.), *Across boundaries. International perspectives on translation studies* (pp. 128–142). Newcastle: Cambridge Scholars Publishing.
- Kelly, D. (2007). Translator competence contextualized. Translator training in the framework of higher education reform: In search of alignment in curricular design. In D. Kenny & K. Ryou (Eds.), *Across boundaries: International perspectives on translation studies* (pp. 128–142). Cambridge, UK: Cambridge Scholars Publishing.
- Kelly, D. (2007). Translator competence contextualized: Translator training in the framework of higher education reform: in search of alignment in curricular design. In D. Kenny & K. Ryou (Eds.), *Across boundaries: International perspectives on translation studies* (pp. 128–142). Cambridge, UK: Cambridge Scholars Publishing.
- Kember, D., Leung, D., Jones, A., Loke, A., McKay, J., & Harrison, T. (2000). Development of a questionnaire to measure the level of reflective thinking. *Assessment & Evaluation in Higher Education*, 25(4), 370–380. doi:10.1080/713611442

- Kenny, M. A. (2008). Discussion, cooperation, collaboration: The impact of task structure on student interaction in a Web-based translation exercise module. *The Interpreter and Translator Trainer*, 2(2), 139–16. doi:10.1080/1750399X.2008.10798771
- Kim, M. (2009). Meaning-oriented assessment of translations. In C. V. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies* (pp. 123–157). Amsterdam: John Benjamins. doi:10.1075/ata.xiv.08kim
- Kim, M. (2013). Research on translator and interpreter education. In C. Millán & F. Bartrina (Eds.), *The Routledge handbook of translation studies* (pp. 102–116). London: Routledge.
- Kiraly, D. (2001). Towards a constructivist approach to translator education. *Quaderns: Revista de traducció*, 6, 50-53.
- Kiraly, D., & Piotrowska, M. (2014). *Towards an emergent curriculum development model for the European Graduate Placement Scheme*. Paper presented at the Future of Education, Florence, Italy.
- Kiraly, D., Rüth, L., Canfora, C., & Cnyrim, A. (Eds.). (2016). *Towards authentic experiential learning in translator education*. Göttingen: V&R unipress/Mainz University Press.
- Kiraly, D. (1995). *Pathways to translation: Pedagogy and process*. Kent, OH: The Kent State University Press.
- Kiraly, D. (1997). Think-aloud protocols and the construction of a professional translator's self-concept. In J. H. Danks, G. M. Shreve, S. B. Fountain, & M. K. McBeath (Eds.), *Cognitive processes in translation and interpreting* (pp. 137–160). Thousand Oaks, CA: Sage.
- Kiraly, D. (2000). *A social constructivist approach to translator education: Empowerment from theory to practice*. Manchester, UK: St Jerome Publishing.
- Kiraly, D. (2003). From Teacher-Centred to Learning-Centred Classrooms in Translator Education: Control, Chaos or Collaboration? In A. Pym, C. Fallada, J. R. Biau, & J. Orenstein (Eds.), *Innovation and E-Learning in Translator Training: Reports on Online Symposia* (pp. 27–31). Tarragona: Universitat Rovira I Virgili.
- Kiraly, D. (2005). Situating praxis in translator education. In K. Károly & Á. Fóris (Eds.), *New trends in translation studies: In honour of Kinga Klauďy* (pp. 117–138). Budapest: Akadémiai Kiadó.
- Kiraly, D. (2013). Towards a view of translation competence as an emergent phenomenon: Thinking outside the box(es) in translator education. In D. Kiraly, S. Hansen-Schirra, & K. Maksymski (Eds.), *New prospects and perspectives for educating language mediators* (pp. 197–224). Tübingen: Gunter Narr Verlag.
- Kiraly, D. (2013). Towards a view of translator competence as an emergent phenomenon: Thinking outside the box (es) in translator education. In D. Kiraly, S. Hansen-Schirra, & K. Maksymski (Eds.), *New prospects and perspectives for educating language mediators* (pp. 197–224). Tübingen: Gunter Narr Verlag.
- Kiraly, D. C. (2003). From instruction to collaborative construction: A passing fad or the promise of a paradigm shift in translator education? In B. J. Baer & G. S. Koby (Eds.), *Beyond the ivory tower: Rethinking translation pedagogy* (pp. 3–27). Amsterdam: John Benjamins. doi:10.1075/ata.xii.03kir
- Klimkowski, K. (2015). *Towards a shared curriculum in translator and interpreter education*. Wrocław: Wydawnictwo Wyższej Szkoły Filologicznej we Wrocławiu, Komisja Nauk Filologicznych PAN, Oddział we Wrocławiu & International Communicology Institute.
- Klimkowski, K. (2006). Team work as a method in translation. *Across Languages and Cultures*, 7(1), 93–103. doi:10.1556/Acr.7.2006.1.6

Compilation of References

- Klimkowski, K. (2016). Entrepreneurial aspects of translation competence in didactic reflection and practice. In Ł. Bogucki, B. Lewandowska-Tomaszczyk, & M. Thelen (Eds.), *Translation and Meaning: New series*, 2 (pp. 143–156). Frankfurt am Main: Peter Lang.
- Klimkowski, K. (2017). Empowering translation students in entrepreneurship: The concept of translation service provision competence in didactic application. In S. Hagemann, J. Neu, & S. Walter (Eds.), *Translationslehre und Bologna-Prozess: Unterwegs zwischen Einheit und Vielfalt* (pp. 189–213). Berlin: Frank & Timme.
- Klimkowski, K., & Klimkowska, K. (2012). Towards empowerment in translator education: Students' opinions and expectations of a translation training course. In S. Hubscher-Davidson & M. Borodo (Eds.), *Global trends in translator and interpreter training: Mediation and Culture* (pp. 180–194). London: Continuum.
- Koby, G. S., Fields, P., Hague, D., Lommel, A., & Melby, A. (2014). Defining translation quality. *Revista Tradumàtica: Tecnologies de la traducció*, 12, 413–420. Retrieved June 13, 2017, from <http://revistes.uab.cat/tradumatica/article/view/n12-koby-fields-hague-et-al/pdf>
- Koby, G., Fields, P., Hague, D., Lommel, A., & Melby, A. K. (2014). Defining translation quality. *Revista Tradumàtica: Tecnologies de la traducció*, 12, 413–420.
- Koby, G. S., & Melby, A. K. (2013). Certification and job task analysis (JTA): Establishing validity of translator certification examinations. *The International Journal of Translation and Interpreting Research*, 5(1), 174–210. doi:10.12807/ti.105201.2013.a10
- Kockaert, H., & Segers, W. (2017). Evaluation of legal translations: PIE method (Preselected Items Evaluation). *JoSTrans, The Journal of Specialised Translation*, 27, 148–163. Retrieved from <http://www.jostrans.org/>
- Kockaert, H., & Segers, W. (2017). Evaluation of legal translations: PIE method (Preselected Items Evaluation). *JoSTrans, The Journal of Specialised Translation*, 27, 148–163. Retrieved July 12, 2017, from <http://www.jostrans.org/>
- Kockaert, H., & Segers, W. (2012). L'assurance qualité des traductions: Items sélectionnés et évaluation assistée par ordinateur. *Meta: Journal des Traducteurs/Meta: Translators' Journal*, 57(1), 159–176.
- Kockaert, H., & Segers, W. (2014). Evaluation de la traduction: La méthode PIE (Preselected Items Evaluation). *Turjuman*, 23(2), 232–250.
- Kockaert, H., & Segers, W. (2017). Evaluation of legal translations: PIE method (Preselected Items Evaluation). *Journal of Specialised Translation*, 27, 148–163.
- Koehn, P. (2017). *Statistical machine translation. 2nd Public Draft*. Retrieved November 17, 2017, from <https://arxiv.org/pdf/1709.07809.pdf>
- Kohn, K., & Kalina, S. (1996). The strategic dimension of interpreting. *Meta: Journal des traducteurs/Meta: Translators' Journal*, 41(1), 118–138.
- Koponen, M. (2016). Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *Journal of Specialised Translation*, 25, 131–148.
- Kostiainen, T. (2012, October). *Tools for quality assurance and translation memory maintenance*. Paper presented at ATA Annual Conference, San Diego, CA.
- Kranich, S., Becher, V., Höder, S., & House, J. (Eds.). (2011). *Multilingual discourse production: Diachronic and synchronic perspectives*. Amsterdam: Benjamins. doi:10.1075/hsm.12
- Krest, M. (1988). Monitoring student writing: How not to avoid the draft. *Journal of Teaching Writing*, 7(1), 27–39.

- Krings, H. P. (1986). Translation problems and translation strategies of advanced German learners of French (L2). In J. House & S. Blum-Kulka (Eds.), *Interlingual and intercultural communication* (pp. 263–276). Tübingen: Gunter Narr Verlag.
- Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research*, 47(2), 211–232. doi:10.3102/00346543047002211
- Künzli, A. (2005). What principles guide translation revision? A combined product and process study. In I. Kemble (Ed.), *Translation norms: What is 'Norma' in the translation profession? Proceedings of the Conference Held on 13th November 2004 in Portsmouth* (pp. 31–43). Portsmouth, UK: University of Portsmouth, School of Languages and Area Studies.
- Künzli, A. (2014). Die Übersetzungsrevision: Begriffsklärungen, Forschungsstand, Forschungsdesiderate. *trans-kom*, 7(1), 1–29. Retrieved from http://www.trans-kom.eu/bd07nr01/trans-kom_07_01_01_Kuenzli_Revision.20140606.pdf
- Künzli, A. (2006). Translation revision: A study of the performance of ten professional translators revising a technical text. In M. Gotti & S. Šarčević (Eds.), *Insights into Specialized Translation* (pp. 195–214). Bern: Peter Lang.
- Künzli, A. (2007). The ethical dimension of translation revision. An empirical study. *The Journal of Specialised Translation*, 8, 42–56.
- Künzli, A. (2009). Qualität in der Übersetzungsrevision: Eine empirische Studie. In H. Kalverkämper & L. Schippel (Eds.), *Translation zwischen Text und Welt: Translationswissenschaft als historische Disziplin zwischen Moderne und Zukunft* (pp. 291–303). Berlin: Frank & Timme.
- Kurz, I. (1989). Conference interpreting: User expectations. In D. L. Hammond (Ed.), *Coming of age*. Proceedings of the 30th annual conference of the American translators association (pp. 143–148). Medford, NJ: Learned Information.
- Kurz, I., & Pöschhacker, F. (1995). Quality in TV interpreting. *Translatio-Nouvelles de la FIT – FIT Newsletter*, 14(3-4), 350–358.
- Kurz, C. (2009). Translatorisches Qualitätsmanagement als verantwortungsvolles Handeln. *Lebende Sprachen*, 54(4), 146–155. doi:10.1515/les.2009.038
- Kurz, I. (1993). Conference interpretation: Expectations of different user groups. *The Interpreter's Newsletter*, 5, 13–21.
- Kurz, I. (1993/2002). Conference interpretation: Expectations of different user groups. *The Interpreters' Newsletter*, 3, 13–21.
- Kurz, I. (2001). Conference interpreting: Quality in the ears of the user. *Meta: Journal des traducteurs Meta:/Translators' Journal*, 46(2), 394–409.
- Kussmaul, P. (1995). *Training the translator*. Amsterdam: John Benjamins. doi:10.1075/btl.10
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). {lmerTest} Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1–26. doi:10.18637/jss.v082.i13
- Lachat, C. (2003). *Problemas y estrategias de traducción* (Unpublished doctoral dissertation). University of Granada, Granada, Spain. Retrieved October 27, 2017, from <http://www.tdx.cat/handle/10803/18013>
- Lai, Y. (2010). Which do students prefer to evaluate their essays: Peers or computer program. *British Journal of Educational Technology*, 41(3), 432–454. doi:10.1111/j.1467-8535.2009.00959.x
- Lång, J., Mäkisalo, J., Gowases, T., & Pietinen, S. (2013). Using eye tracking to study the effect of badly synchronized subtitles on the gaze paths of television viewers. *New Voices in Translation Studies*, 10, 72–86.

Compilation of References

- Larson, M. (1984/1998). *Meaning-based translation: A guide to cross-language equivalence*. Lanham, MD: University Press of America.
- Läubli, S., Fishel, M., Massey, G., Ehrensberger-Dow, M., & Volk, M. (2013). Assessing post-editing efficiency in a realistic translation environment. In *Proceedings of the XIV MT summit workshop on post-editing technology and practice* (pp. 83–91). Allschwil: European Association for Machine Translation.
- Lederer, M. (2007). Can Theory Help Interpreter and Translator Trainers and Trainees? *The Interpreter and Translator Trainer*, 1(1), 15–36. doi:10.1080/1750399X.2007.10798748
- Lee, Y.-H. (2011). Comparing self-assessment and teacher's assessment in interpreter training. *T&I Review*, 1, 87-111.
- Lee, J. (2008). Rating scales for interpreting performance assessment. *The Interpreter and Translator Trainer*, 2(2), 165–184. doi:10.1080/1750399X.2008.10798772
- Lee-Jahnke, H. (2005). New Cognitive approaches in Process-Oriented Translation Training. *Meta*, 50(2), 359–378. doi:10.7202/010942ar
- Lee, S.-B. (2014). An interpreting self-efficacy (ISE) scale for undergraduate students majoring in consecutive interpreting: Construction and preliminary validation. *The Interpreter and Translator Trainer*, 8(2), 183–203. doi:10.1080/1750399X.2014.929372
- Lee, S.-B. (2015). Developing an analytic scale for assessment undergraduate students' consecutive interpreting performances. *Interpreting*, 17(2), 226–254. doi:10.1075/intp.17.2.04lee
- Lee, S.-B. (2016). University students' experience of 'scale-referenced' peer assessment for a consecutive interpreting examination. *Assessment & Evaluation in Higher Education*, 42(7), 1015–1029. doi:10.1080/02602938.2016.1223269
- Lee, S.-B. (2016). University students' experiences of 'scale-referenced' peer assessment for a consecutive interpreting examination. *Assessment & Evaluation in Higher Education*.
- Lee, V. (2015). A model for using the reflective learning journal in the postgraduate translation practice classroom. *Perspectives*, 23(3), 489–505. doi:10.1080/0907676X.2014.948884
- Lee, Y.-H. (2005). Self-assessment as an autonomous learning tool in an interpretation classroom. *Meta*, 50(4). doi:10.7202/019869ar
- Leijen, D. (2014). Applying machine learning techniques to investigate the influence of peer feedback on the writing process. In D. Knorr, C. Heine, & J. Engberg (Eds.), *Methods in writing process research* (pp. 167–183). Frankfurt am Main: Peter Lang.
- Lei, P.-W., & Wu, Q. (2007). CTTITEM: SAS macro and SPSS syntax for classical item analysis. *Behavior Research Methods*, 39(3), 527–530. doi:10.3758/BF03193021 PMID:17958163
- Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence (Cambridge, Mass.)*, 10(3), 282–297. doi:10.1162/105474601300343612
- Levý, J. (2000). Translation as a decision process. In V. Lawrence (Ed.), *The translation studies reader* (pp. 148–159). London: Routledge.
- Li, L., Steckelberg, A. L., & Srinivasan, S. (2008). Utilizing peer interactions to promote learning through a web-based peer assessment system. *Canadian Journal of Learning and Technology (CJLT)*, 34(2). retrieved July 16 2017 from <https://www.cjlt.ca/index.php/cjlt/article/view/26423/19605>

- Liddicoat, A. J. (2016). Intercultural mediation, intercultural communication and translation. *Perspectives*, 24(3), 354–364. doi:10.1080/0907676X.2014.980279
- Lim, H.-O. (2013). A jury of your peers: Peer assessment in consecutive interpretation pedagogy. *Forum*, 11(1), 77–97. doi:10.1075/forum.11.1.04lim
- Lindgren, E., Sullivan, K. P. H., Deutschmann, M., & Steinvall, A. (2009). Supporting learner reflection in the language translation class. *International Journal of Information Communication Technologies and Human Development*, 1(3), 26–48. doi:10.4018/jicthd.2009070102
- Linnankylä, P. (2001). Portfolio: Integrating writing, learning and assessment. In P. Tynjälä, L. Mason, & K. Lonka (Eds.), *Studies in writing: Writing as a learning tool: Integrating theory and practice* (Vol. 7, pp. 145–160). Dordrecht: Kluwer Academic Publishers. doi:10.1007/978-94-010-0740-5_9
- LISA QA Model 3.1. (2006). Retrieved July 4, 2017, from <http://dssresources.com/news/1558.php>
- Lisaitė, D., Maylath, B., Moustén, B., Valdez, S., Castel-Branco, M., Minacori, P., & Vandepitte, S. (2016). Negotiating meaning at a distance: Peer feedback in electronic learning translation environments in the Trans-Atlantic and Pacific Project. In M. Thelen & B. Lewandowska-Tomaszczyk (Eds.), *Translation and meaning. Part 10. Proceedings of the Maastricht Session of the 6th International Maastricht-Lódz Duo Colloquium on "Translation and Meaning", 21-22 May 2015* (pp. 99-114). Maastricht: Zuyd University, Maastricht School of International Communication, Department of Translation and Interpreting.
- Lisaitė, D., Maylath, B., Moustén, B., Valdez, S., Castel-Branco, M., Minacori, P., & Vandepitte, S. (2016). Negotiating meaning at a distance: Peer feedback in electronic learning translation environments in the Trans-Atlantic and Pacific Project. In M. Thelen & B. Lewandowska-Tomaszczyk (Eds.), *Translation and Meaning. Part 10. Proceedings of the Maastricht Session of the 6th International Maastricht-Lódz Duo Colloquium on "Translation and Meaning", 21-22 May 2015* (pp. 99-114). Maastricht: Zuyd University, Maastricht School of International Communication, Department of Translation and Interpreting.
- Lisaitė, D., Vandepitte, S., Maylath, B., Moustén, B., Valdez, S., Castel-Branco, M., & Minacori, P. (2016). Negotiating meaning at a distance: Peer review in electronic learning translation environments. In M. Thelen, G. van Egdom, D. Verbeeck, Ł. Bogucki, & B. Lewandowska-Tomaszczyk (Eds.), *Translation and meaning: New series*, 41(1) (pp. 99–113). Frankfurt am Main: Peter Lang.
- Little, D. (1991). *Learner Autonomy I: Definitions, Issues and Problems*. Dublin: Authentik.
- Littlewood, W. (1996). "Autonomy": An anatomy and a framework. *System*, 24(4), 427–435. doi:10.1016/S0346-251X(96)00039-5
- Littlewood, W. (1997). Self-access: Why do we want it and what can it do? In P. Benson & P. Voller (Eds.), *Autonomy and independence in language learning* (pp. 79–91). London: Longman.
- Liu, M. (2013). Design and analysis of Taiwan's Interpretation Certification Examination. In D. Tsagari & R. van Deemter (Eds.), *Assessment issues in language translation and interpreting* (pp. 163-179). Frankfurt am Main: Peter Lang. Retrieved from <http://orcit.eu>
- Livbjerg, I., & Mees, I. M. (2003). Patterns of dictionary use in non-domain-specific translation. In F. Alves (Ed.), *Triangulating translation: Perspectives in process oriented research* (pp. 123–136). Amsterdam: John Benjamins. doi:10.1075/btl.45.1liv

Compilation of References

- Llewellyn, J. P. (1981). Simultaneous interpreting. In J. K. Woll & M. Deuchar (Eds.), *Perspectives on British Sign Language and Deafness* (pp. 89–104). London: Croom Helm.
- Localization Industry Standards Association (2004). *LISA QA model 3.0. product documentation*. Féchy: LISA.
- Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2), 1–39.
- Lommel, A. (Ed.). (2014). *Multidimensional quality metrics (MQM) definition*. Retrieved June 20, 2017 from <http://www.qt21.eu/mqm-definition/definition-2014-08-19.html>
- Lommel, A., Burchardt, A., & Uszkoreit, H. (2014). Multidimensional quality metrics MQM: A framework for declaring and describing translation quality metrics. *Tradumatica*, 12(12), 455–463. doi:10.5565/rev/tradumatica.77
- Lunt, T., & Curran, J. (2010). 'Are you listening please?' The advantages of electronic audio feedback compared to written feedback. *Assessment & Evaluation in Higher Education*, 35(7), 759–769. doi:10.1080/02602930902977772
- Macfarlane-Dick, D., & Nicol, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. doi:10.1080/03075070600572090
- Malkiel, B. (2009). From Antonia to my Antonia: Tracking self-corrections with Translog. In S. Göpferich, A. L. Jakobsen, & I. M. Mees (Eds.), *Behind the mind: Methods, models and results in translation process research*. Frederiksberg: Samfundslitteratur.
- Mangelsdorf, K. (1992). Peer reviews in the ESL composition classroom: What do the students think? *ELT Journal*, 46(3), 274–284. doi:10.1093/elt/46.3.274
- Mangelsdorf, K., & Schlumberger, A. (1992). ESL student response stances in a peer-review task. *Journal of Second Language Writing*, 1(3), 235–254. doi:10.1016/1060-3743(92)90005-A
- Margolies, E. (2015). Going to hear a dance: On audio describing. *Performance Research*, 2(6), 17–23. doi:10.1080/13528165.2015.1111044
- Mariana, V., Cox, T., & Melby, A. (2015). The Multidimensional Quality Metrics (MQM) Framework: A new framework for translation quality assessment. *The Journal of Specialised Translation*, 23, 137–161.
- Mariana, V., Cox, T., & Melby, A. K. (2015). The multidimensional quality metrics (MQM) framework: A new framework for translation quality assessment. *The Journal of Specialised Translation*, 23, 137–161.
- Marrone, S. (1993). Quality: A shared objective. *The Interpreters' Newsletter*, 5, 35–41.
- Martínez Melis, N., & Hurtado Albir, A. (2001). Assessment in translation studies: Research needs. *Meta*, 46(2), 272–287. doi:10.7202/003624ar
- Martínez Melis, N., & Hurtado Albir, A. (2001). Assessment in translation studies: Research needs. *Meta: Journal des Traducteurs / Meta: Translators' Journal*, 46(2), 272–287.
- Martínez Sierra, J. J. (2009). The challenge of translating humour for dubbing. Some problematical issues. In A. Goldstein & B. Golubović (Eds.), *Foreign language movies – Dubbing vs. subtitling* (pp. 129–149). Hamburg, Germany: Verlag Dr. Kovač.
- Martínez-Melis, N., & Hurtado Albir, A. (2001). Assessment in translation studies: Research needs. *Meta, Journal des Traducteurs/Meta: Translators' Journal*, 46(2), 272–287.

- Marzà Ibañez, A. (2010). Evaluation criteria and film narrative. A frame to teaching relevance in audio description. *Perspectives: Studies in Translatology*, 18(3), 143–153. doi:10.1080/0907676X.2010.485682
- Massardo, I., Van der Meer, J., & Khalilov, M. (2016). *TAUS translation technology landscape report*. De Rijp: TAUS.
- Massardo, I., van der Meer, J., & Khalilov, M. (2016). *Translation technology landscape report. September 2016*. De Rijp: TAUS.
- Massey, G. (2005). Process-oriented translator training and the challenge for E-learning. *Meta*, 50(2), 626–633. doi:10.7202/011006ar
- Massey, G., & Ehrensberger-Dow, M. (2011). Commenting on translation: Implications for translator training. *The Journal of Specialised Translation*, 16, 26–41.
- Massey, G., & Ehrensberger-Dow, M. (2011a). Commenting on translation: Implications for translator training. *Journal of Specialised Translation*, 16, 26–41.
- Massey, G., & Ehrensberger-Dow, M. (2011b). Investigating information literacy: A growing priority in translation studies. *Across Languages and Cultures*, 12(2), 193–211. doi:10.1556/Acr.12.2011.2.4
- Massey, G., & Ehrensberger-Dow, M. (2012). Evaluating the process: Implications for curriculum development. In L. Zybatow, A. Petrova, & M. Ustaszewski (Eds.), *Translationwissenschaft interdisziplinär: Fragen der Theorie und der Didaktik* (pp. 95–100). Frankfurt am Main: Peter Lang.
- Massey, G., & Ehrensberger-Dow, M. (2013). Evaluating translation processes: Opportunities and challenges. In D. Kiraly, S. Hansen-Schirra, & K. Maksymski (Eds.), *New prospects and perspectives for educating language mediators* (pp. 157–180). Tübingen: Gunther Narr.
- Massey, G., & Ehrensberger-Dow, M. (2014). Looking beyond text. The usefulness of translation process data. In D. Knort, C. Heine, & J. Engberg (Eds.), *Methods in writing process research* (pp. 81–98). Frankfurt am Main: Peter Lang.
- Mast, C. (2016). *Unternehmenskommunikation. Ein Leitfadens* (6th ed.). Konstanz: UVK.
- Matamala, A., & Remael, A. (2015). Audio-description reloaded: An analysis of visual scenes in 2012 and Hero. *Translation Studies*, 8(1), 63–81. doi:10.1080/14781700.2014.943678
- Matlock-Hetzel, S. (1997, January). *Basic concepts in item and test analysis*. Paper presented at Annual meeting of the Southwest Educational Research Association, Austin, TX.
- Matsumoto, K. (1996). Helping 12 learners reflect on classroom learning. *ELT Journal*, 50(2), 143–149. doi:10.1093/elt/50.2.143
- Maylath, B. (1997). Writing globally: Teaching the technical communication student to prepare documents for translation. *Journal of Business and Technical Communication*, 11(3), 339–352. doi:10.1177/1050651997011003006
- Mayoral, R. (2001). El espectador y la traducción audiovisual. In F. Chaume & R. Agost (Eds.), *La traducción en los medios audiovisuales* (pp. 33–46). Castellón de la Plana, Spain: Publicacions de la Universitat Jaume I.
- Mayoral, R. (2005). Reflexiones sobre la investigación en traducción audiovisual. In P. Zabalbeascoa, L. Santamaría, & F. Chaume (Eds.), *La traducción audiovisual: Investigación, enseñanza y profesión* (pp. 3–8). Granada, Spain: Comares.
- McAlester, G. (2000). The evaluation of translation into a foreign language. In C. Schäffner & B. Adab (Eds.), *Developing translation competence* (pp. 229–241). Amsterdam: John Benjamins. doi:10.1075/btl.38.21mca

Compilation of References

- McClarty, K. L., & Gaertner, M. N. (2015). *Measuring mastery. Best practices for assessment in competency-based education*. Retrieved from <http://www.aei.org/publication/measuring-mastery-best-practices-for-assessment-in-competency-based-education/>
- McMillan, K., & Weyers, J. (2011). *How to write dissertations and project reports* (2nd ed.). Harlow, UK: Pearson Education.
- McMillan, L. (2010). Students write to students about writing. In J. Harris, J. D. Miles, & C. Paine (Eds.), *Teaching with student texts: Essays toward an informed practice* (pp. 88–95). Logan, UT: Utah State University Press. doi:10.2307/j.ctt4cgm3.13
- Mead, P. (2005). Methodological issues in the study of interpreters' fluency. *The Interpreters' Newsletter*, 13, 39–63.
- Melby, A., Fields, P., Hague, D. Koby, G., & Lommel, A. (2014). Defining the landscape of translation. *Revista Tradumàtica: tecnologies de la traducció*, 12, 392–403.
- Mellinger, C. D., & Shreve, G. M. (2016). Match evaluation and over-editing in a translation memory environment. In R. Muñoz Martín (Ed.), *Reembedding translation process research* (pp. 131–148). Amsterdam: Benjamins. doi:10.1075/btl.128.07mel
- Melton, J. H. (2008). Lost in translation: Professional communication competencies in global training contexts. *IEEE Transactions on Professional Communication*, 52(2), 198–214. doi:10.1109/TPC.2008.2000346
- Mertin, E. (2006). *Prozessorientiertes Qualitätsmanagement im Dienstleistungsbereich Übersetzen. Leipziger Studien Zur Angewandten Linguistik Und Translatologie*, 2. Frankfurt am Main: Peter Lang.
- Mezirow, J. (1991). *Transformative dimensions of adult learning*. San Francisco: Jossey-Bass.
- Miguélez, C. (1999, Spring). Current issues in court interpreting: Spain, a case study. *Proteus: Newsletter of the National Association of Judiciary Interpreters and Translators*, 8-2. Retrieved from http://www.najit.org/members_only/proteus/v8n2/miguelezv8n2.htm
- Mikkelsen, H. (2013). Universities and interpreter certification. *The International Journal of Translation and Interpreting Research*, 5(1), 66–78. doi:10.12807/ti.105201.2013.a03
- Min, H.-T. (2005). Training students to become successful peer reviewers. *System*, 33(2), 293–308. doi:10.1016/j.system.2004.11.003
- Minsky, M. (1980). Telepresence. *Omni (New York, N.Y.)*, 2, 45–51.
- Miquel, M. (2014). The reception of subtitling by the deaf and hard of hearing. Preliminary findings. In E. Torres-Simon & D. Orrego-Carmona (Eds.), *From translation research projects 5* (pp. 63–76). Tarragona, Spain: Intercultural Studies Group.
- Mobaraki, M., & Aminzadeh, S. (2012). A study on different translation evaluation strategies to introduce an eclectic method. *International Journal of English Linguistics*, 2(6), 63–70. doi:10.5539/ijel.v2n6p63
- Moon, J. A. (1999). *Reflection in learning and professional development*. London: RoutledgeFalmer.
- Moon, J. A. (2004). *A handbook of reflective and experiential learning: Theory and practice*. London: RoutledgeFalmer.
- Moser-Mercer, B. (1996). Quality in interpreting: Some methodological issues. *The Interpreters' Newsletter*, 7, 43–56.
- Mossop, B. (2007). Empirical studies of revision: What we know and need to know. *JosTrans: The Journal of Specialised Translation*, 8, 5–20.

- Mossop, B. (2014). *Revising and editing for translators* (3rd ed.). New York: Routledge.
- Munday, J. (2008). *Introducing Translation Studies: Theories and applications*. London: Routledge Taylor & Francis.
- Muñoz Martín, R. (2009). Typos & Co. In S. Göpferich, A. Jakobsen, & I. Mees (Eds.), *Behind the Mind: Methods, Models and Results in Translation Process Research* (pp. 167–191). Copenhagen: Samfundslitteratur.
- Muñoz, R., & Conde, T. (2007). Effects of serial translation evaluation. In P. A. Schmidt & H. E. Jüngst (Eds.), *Translationsqualität* (pp. 428–444). Frankfurt, Germany: Peter Lang.
- Murdock, J. (2014). Perturbation models. In M. Grinfield (Ed.), *Mathematical tools for physicists* (pp. 411–449). Somerset, NJ: Wiley.
- Murray, D. M. (1972). Teach writing as a process not product. *The Leaflet*, 11-14.
- Murray, D. M. (1968). *A writer teaches writing: A practical method of teaching composition*. Boston: Houghton Mifflin.
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142. doi:10.1111/j.2041-210x.2012.00261.x
- National Accreditation Authority for Translators and Interpreters (NAATI). (2015). *NAATI translator certification: Knowledge, skills and attributes*. Retrieved June 3, 2017, from https://www.naati.com.au/media/1059/translator_ksa_paper.pdf
- National Center for State Courts – Language Access Services Section. (2014). *Court interpreter oral examination: Test Construction Manual*. Retrieved from http://www.ncsc.org/~media/Files/PDF/Services%20and%20Experts/Areas%20of%20expertise/Language%20Access/Written%20and%20Oral/2014%20January_Test%20Construction%20Manual%201%2029%2014.ashx
- National Center for State Courts. (2006). *Court interpreting qualifications*. Retrieved from <http://www.ncsc.org/~media/Files/PDF/Education%20and%20Careers/State%20Interpreter%20Certification%202/Qualifications.ashx>
- National Center for State Courts. (2016, November 9). *Court interpreter oral exams. NCSC oral examinations ready for administration (20 languages)*. Retrieved from http://www.ncsc.org/~media/Files/PDF/Services%20and%20Experts/Areas%20of%20expertise/Language%20Access/Resources%20for%20Program%20Managers/Oral_Exam_Ready_for_Administration%20November%202016.ashx
- National Center for State Courts. (2017, September 19). *NCSC oral examinations ready for administration*. Retrieved from http://www.ncsc.org/~media/files/pdf/services%20and%20experts/areas%20of%20expertise/language%20access/written%20and%20oral/oral_exam_ready_for_administration%20september%202017.ashx
- Nelson, G. L., & Carson, J. G. (1998). ESL students' perceptions of effectiveness in peer response groups. *Journal of Second Language Writing*, 7(2), 113–131. doi:10.1016/S1060-3743(98)90010-8
- Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science*, 37(4), 375–401. doi:10.1007/11251-008-9053-x
- Neubert, A. (1994). Competence in translation: A complex skill, how to study and how to teach it. In M. Snell-Hornby, F. Pöchhacker, & K. Kaindl (Eds.), *Translation studies: An interdiscipline* (pp. 411–420). Amsterdam: John Benjamins. doi:10.1075/btl.2.48neu
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.
- Newmark, P. (1988). *A textbook of translation*. New York: Prentice Hall.

Compilation of References

- Ng, B. C. (1992). End users' subjective reaction to the performance of student interpreters. *The Interpreters' Newsletter*, 1(Special Issue), 35–41.
- Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*, 35(5), 501–517. doi:10.1080/02602931003786559
- Nicol, D. (2011). *Developing the students' ability to construct feedback*. Gloucester, UK: Quality Assurance Agency for Higher Education.
- Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education*, 39(1), 102–122. doi:10.1080/02602938.2013.795518
- Nida, E., & Taber, C. (1969). *The theory and practice of translation*. Leiden: Brill.
- Niska, H. (2002). Community interpreter training. Past, present, future. In G. Garzone & M. Viezzi (Eds.), *Interpreting in the 21st century* (pp. 133–144). Amsterdam: John Benjamins. doi:10.1075/btl.43.14nis
- Nonaka, I. (1991, November). The knowledge-creating company. *Harvard Business Review*, 96–104.
- Nord, C. (2002). *Lernziel: Professionelles Übersetzen: Spanish-Deutsch*. Wilhelmsfeld: Gottfried Egert.
- Nord, C. (2011b). *Defining translation functions: The translation brief as a guideline for the trainee translator*. Retrieved June 20, 2017 from <http://www.scribd.com/doc/39090525/NORD>
- Nord, C. (1988). *Textanalyse und Übersetzen: Theoretische Grundlagen, Methode und didaktische Anwendung einer übersetzungsrelevanten Textanalyse*. Heidelberg, Germany: Groos.
- Nord, C. (1991). *Text analysis in translation: Theory, methodology, and didactic application of a model for translation-oriented text analysis*. Amsterdam: Rodopi.
- Nord, C. (1992). Text analysis in translator training. In C. Dollerup & A. Loddegaard (Eds.), *Teaching translation and interpreting: Training, talent, and experience* (pp. 39–48). Amsterdam: John Benjamins. doi:10.1075/z.56.08nor
- Nord, C. (1994). It's tea-time in wonderland: 'Culture-markers' in fictional texts. In H. Pürschel (Ed.), *Intercultural communication* (pp. 523–538). Frankfurt: Peter Lang.
- Nord, C. (1997). *Translating as a purposeful activity: Functionalist approaches explained*. Manchester, UK: St. Jerome.
- Nord, C. (2000). Loyalty revisited: Bible translation as a case in point. *The Translator*, 7(2), 195–202.
- Nord, C. (2010). Functionalist approaches. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of Translation Studies, 1* (pp. 120–128). Amsterdam: John Benjamins. doi:10.1075/hts.1.fun1
- Nord, C. (2011a). *Funktionsgerechtigkeit und Loyalität: Theorie, Methode und Didaktik des funktionalen Übersetzens*. Berlin: Frank & Timme.
- O'Brien, S. (2011). Collaborative translation. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of translation studies, 2* (pp. 17–20). Amsterdam: John Benjamins. doi:10.1075/hts.2.col1
- O'Brien, S. (2012). Towards a dynamic quality evaluation model for translation. *Journal of Specialized Translation*, 17, 55–77.
- O'Brien, S. (2012). Towards a Dynamic Quality Evaluation Model for Translation. *The Journal of Specialised Translation*, 17, 55–77.
- O'Brien, S. (2012). Translation as human-computer interaction. *Translation Spaces, 1*, 101–122. doi:10.1075/ts.1.05obr

- O'Brien, S., Winther Balling, L., Carl, M., Simard, M., & Specia, L. (Eds.). (2014). *Post-editing of machine translation: Processes and applications*. Newcastle: Cambridge Scholars.
- Olohan, M. (2011). Translators and translation technology: The dance of agency. *Translation Studies*, 4(3), 342–357. doi:10.1080/14781700.2011.589656
- Olvera Lobo, M. D., Robinson, B., Castro Prieto, R. M., Quero Gervilla, E., Muñoz Raya, E., Murillo Melero, M., ... Díez Lerma, J. L. (2007). A professional approach to translator training (PATT). *Meta*, 52(3), 517–528. doi:10.7202/016736ar
- Op de Beeck, J. (2015). *Het Verlies van België*. Antwerpen: Horizon.
- OPTIMALE WP 5.4. (2013a). *Quality assurance*. Retrieved from <http://translator-training.eu/training/new-tools-and-technologies/quality-assurance>
- OPTIMALE WP 5.4. (2013b). *Status quo report, WP 5 – Training for professional practice, WP 5.4 – Quality Assurance*. Retrieved from http://www.ressources.univ-rennes2.fr/service-relations-internationales/optimale/attachments/article/15/WP5.4%20Synthesis%20report_EN.pdf
- OPTIMALE WP 5.4. (2013c). *Case studies of good practice in translator training - QA in the translation market*. Retrieved from http://www.ressources.univ-rennes2.fr/service-relations-internationales/optimale/attachments/article/15/WP5.4%20Case%20study_EUATC_EN.pdf
- Orero, P. (2005). Audio description: Professional recognition, practice and standards in Spain. *Translation Watch Quarterly*, 1(Dec), 7–18.
- Orero, P. (2005). La traducción de entrevistas para voice-over. In P. Zabalbeascoa, L. Santamaría, & F. Chaume (Eds.), *La traducción audiovisual: Investigación, enseñanza y profesión* (pp. 213–222). Granada, Spain: Comares.
- Orlando, M. (2011). Evaluation of translations in the training of professional translators: At the crossroads between theoretical, professional and pedagogical practices. *The Interpreter and Translator Trainer*, 5(2), 293-308.
- Orlando, M. (2011). Evaluation of Translations in the Training of Professional Translators: At the Crossroads between theoretical, professional and pedagogical practices. *The Interpreter and Translator Trainer*, 5(2), 293-308.
- Orlando, M. (2011). Evaluation of Translations in the Training of Professional Translators: At the Crossroads between Theoretical, Professional and Pedagogical Practices. *The Interpreter and Translator Trainer*, 5(2), 293-308.
- Orlando, M. (2012). Training of professional translators in Australia: Process-oriented and product oriented evaluation approaches. In S. Hubscher-Davidson & M. Borodo (Eds.), *Global trends in translator and interpreter training: Mediation and culture* (pp. 197–216). London: Continuum.
- Orlando, M. (2012). Training of professional translators in Australia: Process-oriented and product-oriented evaluation approaches. In S. Hubscher-Davidson & M. Borodo (Eds.), *Global trends in translator and interpreter training* (pp. 197–216). London: Continuum.
- Orlando, M. (2012). Training Professional Translators in Australia: Process-oriented and Product-oriented Evaluation Approaches. In S. Hubscher-Davidson & M. Borodo (Eds.), *Global Trends in Translator and Interpreter Training* (pp. 197–216). London: Continuum.
- Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education*, 25(1), 23–38. doi:10.1080/02602930050025006

Compilation of References

- Ortega, J. M., Sandrelli, A., & Townsley, B. (2014). Interpreting test format. In C. Giambruno (Ed.), *Assessing legal interpreter quality through testing and certification: The QUALITAS project* (pp. 41-67). Sant Vicent del Raspeig: Publicaciones de la Universidad de Alicante.
- Ozolins, U. (2016). The long road to effective court interpreting. *Language and the Law / Linguagem E Direito*, 3(2), 98-115.
- PACTE. (2002). Exploratory tests in a study of translation competence. *Conference Interpretation and Translation*, 4(2), 41-69.
- PACTE. (2003). Building a translation competence model. In F. Alves (Ed.), *Triangulating translation* (pp. 43-66). Amsterdam: John Benjamins.
- PACTE. (2003). Building a translation competence model. In F. Alves (Ed.), *Triangulating translation: Perspectives in process oriented research* (pp. 43-66). Amsterdam: John Benjamins.
- PACTE. (2003). Building a translation competence model. In F. Alves (Ed.), *Triangulating translation: Perspectives in process-oriented research* (pp. 43-66). Amsterdam: John Benjamins.
- PACTE. (2005). Investigating translation competence: Conceptual and methodological issues. *Meta*, 50(2), 609-619. doi:10.7202/011004ar
- PACTE. (2005). Primeros resultados de un experimento sobre la competencia traductora. In M. L. Romana (Ed.), *AIETI formación, investigación y profesión: II congreso internacional AIETI, Información y documentación* (pp. 573-587). Madrid: AETI.
- PACTE. (2008). First results of a translation competence experiment: 'Knowledge of translation' and 'efficacy of the translation process. In J. Kearns (Ed.), *Translator and interpreter training: Issues, methods and debates* (pp. 104-126). London: Continuum.
- PACTE. (2011). Results of the validation of the PACTE translation competence model: Translation problems and translation competence. In C. Alvstad, A. Hild, & E. Tiselius (Eds.), *Methods and strategies of process research: Integrative approaches in translation studies* (pp. 317-343). Amsterdam: John Benjamins.
- PACTE. (2016a, July). *Nivelación de competencias en la adquisición de la competencia traductora*. Paper presented at DidTRAD III. Congreso internacional sobre investigación en Didáctica de la traducción, Barcelona, Spain.
- PACTE. (2016b, September). *Competence levels in translation: Working towards a European framework*. Paper presented at the 8th EST Congress, Aarhus.
- PACTE. (2017). The second stage of PACTE group's research: Experimental research into the acquisition of translation competence. In A. Hurtado Albir (Ed.), *Researching Translation Competence by PACTE Group* (pp. 303-308). Amsterdam: John Benjamins.
- Paradigm Testing, L. L. C. (2017, June 1). *Federal Court Interpreter Certification Examination for Spanish/English*. Retrieved from <https://paradigmtesting.com/fcice-hb/>
- Paran, A. (2017). 'Only Connect': Researchers and teachers in dialogue. *ELT Journal*, 71(4), 499-508. doi:10.1093/elt/ccx033
- Paton, F. (2002). Approaches to productive peer review. In D. Roen, V. Pantoja, L. Yena, S. Miller, & E. Waggoner (Eds.), *Strategies for teaching first-year composition* (pp. 301-306). Urbana, IL: NCTE.

- Paulus, T. M. (1999). The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing*, 8(3), 265–289. doi:10.1016/S1060-3743(99)80117-9
- Pavani, S. (2016). *La evaluación sumativa en la didáctica de la traducción: Un estudio empírico en la enseñanza de la traducción español-italiano* (Doctoral dissertation). Retrieved December 28, 2017, from https://ddd.uab.cat/pub/tesis/2016/hdl_10803_400016/stpa1de1.pdf
- Pavlović, N. (2007). *Directionality in collaborative translation processes* (Doctoral dissertation). Retrieved from <http://www.thesisenxarxa.net/TDX-1210107-172129/>
- Pedrola, M. (1999). Interview with Peter Newmark. In G. Anderman & M. Rogers (Eds.), *Word, text, translation: Liber Amicorum for Peter Newmark* (pp. 17–25). Bristol, UK: Multilingual Matters.
- Perrin, D., & Stücheli-Herlach, P. (2013). Schreiben mit System: Texte planen, entwerfen und verbessern für die medienkonvergente PR. In D. Perrin & P. Stücheli-Herlach (Eds.), *Schreiben mit System* (pp. 15–38). Wiesbaden: Springer VS.
- Pöchhacker, F. (1994). Quality assurance in simultaneous interpreting. In C. Dollerup, & A. Lindegaard (Eds.), *Teaching translation and interpreting 2: Insights, aims and visions. Papers from the Second Language International Conference, Elsinore* (pp. 233–242). Amsterdam: John Benjamins.
- Pöchhacker, F. (1993). On the science of interpretation. *The Interpreters' Newsletter*, 5, 52–59.
- Pöchhacker, F. (2001). Quality assessment in conference and community interpreting. *Meta: Journal des traducteurs / Meta: Translators' Journal*, 46(2), 410–425.
- Pöchhacker, F. (2004). I in TS: On partnership in translation studies. In C. Schäffner (Ed.), *Translation research and interpreting research: traditions, gaps and synergies* (pp. 104–115). Clevedon, UK: Multilingual Matters.
- Pöchhacker, F. (2016). *Introducing interpreting studies*. London: Routledge.
- Polanski, R., & Benmussa, R. (Producers) & Polanski, R. (Director). (2002) *The Pianist* [Motion Picture]. UK, Focus Features.
- Popovič, A. (1975). *Teória umeleckého prekladu*. Bratislava: Tatran.
- Popovič, A. (1976). *Dictionary for the analysis of literary translation*. Edmonton: University of Alberta.
- Popovič, A. (2006). *La scienza della traduzione: Aspetti metodologici: La comunicazione traduttiva (B. Osimo & D. Laudani, Trans.)*. Milan: Hoepli.
- Postigo Pinazo, E. (2008). Self-assessment in teaching interpreting. *TTR: traduction, terminologie, redaction*, 21(1), 173–211.
- Postigo Pinazo, E. (2008). Self-Assessment in Teaching Interpreting. *TTR*, 21(1), 173–209. doi:10.7202/029690ar
- Postigo Pinazo, E. (2008). Self-assessment in teaching interpreting. *TTR: Traduction, Terminologie. Redaction*, 21(1), 173–211.
- Pradas Macías, E. M. (2006). Probing quality criteria in simultaneous interpreting: The role of silent pauses. *Interpreting*, 8(1), 25–43. doi:10.1075/intp.8.1.03pra
- Prassl, F. (2010). Translators' decision-making processes in research and knowledge integration. In S. Göpferich, F. Alves, & I. M. Mees (Eds.), *New approaches in translation process research* (pp. 57–81). Copenhagen: Samfundslitteratur Press.

Compilation of References

- Pretz, J. E., Naples, A. J., & Sternberg, R. J. (2003). Recognizing, defining, and representing problems. In J. E. Davidson & R. J. Sternberg (Eds.), *The psychology of problem solving*. New York: Cambridge University Press. doi:10.1017/CBO9780511615771.002
- Price, M., & O'Donovan, B. (2006). Improving student performance through enhanced student understanding of criteria and feedback. In C. Bryan & K. Clegg (Eds.), *Innovative assessment in higher education* (pp. 100–109). London: Routledge.
- Price, M., Rust, C., O'Donovan, B., Hindley, K., & Bryant, R. (2012). *Assessment literacy: The foundation for improving student learning*. Oxford, UK: Oxford Centre for Staff and Learning Development.
- Prieto Ramos, F. (2011). Developing legal translation competence: An integrative process-oriented approach. *Comparative Legilinguistics - International Journal for Legal Communication*, 5, 7-21.
- Prieto Ramos, F. (2014). Parameters for problem-solving in legal translation: Implications for legal lexicography and institutional terminology management. In *The Ashgate handbook of legal translation* (pp. 135–152). Farnham, UK: Ashgate.
- Prieto Ramos, F. (2014). Parameters for problem-solving in legal translation: Implications for legal lexicography and institutional terminology management. In L. Cheng, K. K. Sin, & A. Wagner (Eds.), *The Ashgate handbook of legal translation* (pp. 121–134). London: Routledge.
- Puntambekar, S., & Hubscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist*, 40(1), 1–12. doi:10.120715326985ep4001_1
- Pym, A. (1992). Translation error analysis and the interface with language teaching. In C. Dollerup & A. Loddegaard (Eds.), *Teaching translation and interpreting* (pp. 279–288). Amsterdam: John Benjamins. doi:10.1075/z.56.42pym
- Pym, A. (2003). Redefining translation competence in an electronic age. In defence of a minimalist approach. *Meta*, 48(4), 481–497. doi:10.7202/008533ar
- Pym, A. (2011). Training translators. In K. Malmkjær & K. Windle (Eds.), *The Oxford handbook of translation studies* (pp. 475–489). New York: Oxford University Press.
- Pym, A., Orrego-Carmona, D., & Torres-Simón, E. (2015). Status and technology in the professionalization of translators. Market disorder and the return of hierarchy. *Journal of Specialised Translation*, 25, 33–53.
- Quirke, B. (2008). *Making the connections: Using internal communication to turn strategy into action*. Wallingford: Gower.
- R Development Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: The R Foundation for Statistical Computing.
- Race, P. (2014). *Making learning happen: A guide for post-compulsory education* (3rd ed.). Los Angeles: Sage.
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28(1), 4–13. doi:10.1002/bs.3830280103
- Ramos, M. (2015). The emotional experience of films: Does Audio Description make a difference? *The Translator*, 21(1), 68–94. doi:10.1080/13556509.2014.994853
- Rasinger, S. M. (2013). *Quantitative research in linguistics: An introduction* (2nd ed.). London: Bloomsbury.
- Rasmussen, K. W., & Schjoldager, A. (2011). Revising translations: A survey of revision policies in Danish translation companies. *JosTrans: The Journal of Specialised Translation*, 15, 87–120.
- Remael, A. (2005). *Audio description for recorded TV, cinema and DVD. An experimental stylesheet for teaching purposes*. Retrieved from <http://www.hivt.be>

- Risku, H. (2010). A cognitive scientific view on technical communication and translation: Do embodiment and situatedness really make a difference? *Target*, 22(1), 94–111. doi:10.1075/target.22.1.06ris
- Risku, H., & Dickinson, A. (2009). Translators as networkers: The role of virtual communities. *Hermes*, 42, 49–70.
- RNIB. (n.d.). Retrieved from <https://www.rnib.org.uk/rnibconnect/how-i-see>
- Robert, I. S. (2008). *Translation revision procedures: An explorative study*. Retrieved from <http://www.arts.kuleuven.be/cetra/papers/files/robert.pdf>
- Robert, I. S. (2012). *La révision en traduction: les procédures de révision et leur impact sur le produit et le processus de révision* (Unpublished doctoral dissertation). University of Antwerp, Antwerpen.
- Robert, I. (2012). *La révision en traduction: les procédures de révision et leur impact sur le produit et le processus de révision*. Antwerp: University of Antwerp. Retrieved from http://www.alineremael.be/data/These_ISBN_20120425_BW.pdf
- Robert, I. S. (2014). Investigating the problem-solving strategies of revisers through triangulation: An exploratory study. *Translation and Interpreting Studies*, 9(1), 88–108. doi:10.1075/tis.9.1.05rob
- Robert, I., Remael, A., & Ureel, J. J. J. (2016). Towards a model of translation revision competence. *The Interpreter and Translator Trainer*, 11(1), 1–19. doi:10.1080/1750399X.2016.1198183
- Robertson, S. I. (2001). *Problem solving*. Philadelphia: Psychology Press. doi:10.4324/9780203457955
- Robinson, B., López Rodríguez, C., & Tercedor Sánchez, M. (2006). Self-assessment in translator training. *Perspectives. Studies in Translatology*, 14(2), 115–138. doi:10.1080/09076760608669025
- Rodgers, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers College Record*, 104(4), 842–866. doi:10.1111/1467-9620.00181
- Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT Journal*, 59(1), 23–30. doi:10.1093/elt/cci003
- Romero-Fresco, P. (2013). Accessible filmmaking: Joining the dots between audiovisual translation, accessibility and filmmaking. *The Journal of Specialised Translation*, 20, 201–223.
- Romero-Trillo, J. (2015). *Yearbook of corpus linguistics and pragmatics 2015: Current Approaches to discourse and translation studies*. Cham: Springer. doi:10.1007/978-3-319-17948-3
- Russo, M. (1995). Self-evaluation: The awareness of one's own difficulties as a training tool for simultaneous interpretation. *The Interpreters' Newsletter*, 6, 75–85.
- Russo, M. (2011). Aptitude testing over the years. *Interpreting*, 13(1), 5–30. doi:10.1075/intp.13.1.02rus
- Rust, C., Price, M., & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education*, 28(2), 147–164. doi:10.1080/02602930301671
- Sacks, O. W. (1998). *The man who mistook his wife for a hat and other clinical tales*. London: Gerald Duckworth.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144. doi:10.1007/BF00117714
- Sadler, D. R. (1998). Formative assessment: Revising the territory. *Assessment in Education: Principles, Policy & Practice*, 5(1), 77–84. doi:10.1080/0969595980050104
- SAE International. (2005). *Translation quality metric SAE J2450_201608*. Retrieved July 4, 2017, from http://standards.sae.org/j2450_200508/

Compilation of References

- Saldanha, G., & O'Brien, S. (2014). *Research methodologies in translation studies*. London: Routledge.
- Saldhana, G., & O'Brien, S. (Eds.). (2013). *Research Methodologies in Translation Studies*. London: Routledge.
- Sandrelli, A. (2015). Becoming an interpreter: The role of computer technology. *MonTI*, 2, 111–138.
- Sawyer, D. B. (2004). *Fundamental aspects of interpreter education: Curriculum and assessment*. Amsterdam: John Benjamins. doi:10.1075/btl.47
- Scardamalia, M. (2001). Big change questions: Will educational institutions, within their present structures, be able to adapt sufficiently to meet the needs of the information age? *Journal of Educational Change*, 2(2), 171–176. doi:10.1023/A:1017982102060
- Schaffer, J. (1996). Peer response that works. *Journal of Teaching Writing*, 15(1), 81–89.
- Schäffner, C., & Adab, B. (2000). Developing translation competence: Introduction. In C. Schäffner & B. Adab (Eds.), *Developing translation competence* (Vol. 38, pp. vii–xvi). Amsterdam: John Benjamins. doi:10.1075/btl.38.01sch
- Schäffner, C., & Adab, B. (Eds.). (2000). *Developing translation competence*. Amsterdam: John Benjamins. doi:10.1075/btl.38
- Schiefele, U. (1991). Interest, learning, and motivation. *Educational Psychologist*, 26(3-4), 299–323. doi:10.1080/00461520.1991.9653136
- Schjoldager, A. (1996). Assessment of simultaneous interpreting. In C. Dollerup & V. Appel (Eds.), *Teaching translation and interpreting 3. New horizons: Papers from the third language international conference* (pp. 187–195). Amsterdam: John Benjamins.
- Schjoldager, A. (1996). Assessment of simultaneous interpreting. In C. Dollerup & V. Appel (Eds.), *Teaching Translation and Interpreting 3: New Horizons* (pp. 187–195). Amsterdam: John Benjamins. doi:10.1075/btl.16.26sch
- Schmid, B., & Lyczek, B. (2006). Die Rolle der Kommunikation in der Wertschöpfung der Unternehmung. In B. Schmid & B. Lyczek (Eds.), *Unternehmenskommunikation. Kommunikationsmanagement aus Sicht der Unternehmensführung* (pp. 3–146). Wiesbaden: Gabler. doi:10.1007/978-3-8349-9126-3_1
- Schön, D. A. (1983). *The reflective practitioner*. New York: Basic Books.
- Schön, D. A. (1987). *Educating the reflective practitioner*. San Francisco: Jossey Bass.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (Vol. I, pp. 39–83). Chicago, IL: Rand McNally.
- Segers, W. (2012). Vertaalevaluatie: Waarover hebben we het eigenlijk? *Ad Rem. Tijdschrift voor zakelijke communicatie*, 26(6), 17–18.
- Segers, W., & Kockaert, H. (2016). Can subjectivity be avoided in translation evaluation? In M. Thelen, G.-W. van Egdom, D. Verbeeck, Ł. Bogucki, & B. Lewandowska-Tomaszczyk (Eds.), *Translation and meaning: New series, 1* (pp. 69–78). Frankfurt a/Main: Peter Lang.
- Segers, W., & Kockaert, H. (2016). Can subjectivity be avoided in translation evaluation? In *Translation and meaning: New series, 1* (pp. 69–78). Frankfurt am Main: Peter Lang.
- Segovia, R. (2005). Semiótica textual y traducción audiovisual. In P. Zabalbeascoa, L. Santamaría, & F. Chaume (Eds.), *La traducción audiovisual: Investigación, enseñanza y profesión* (pp. 81–92). Granada, Spain: Comares.
- Shebani, B. (Director). (2017). *Barbershop Chronicles*. National Theatre of Great Britain.

- Shei, C. C.-C. (2005). Translation commentary: A happy medium between translation curriculum and EAP. *System*, 33(2), 309–325. doi:10.1016/j.system.2005.01.004
- Shih, C. (2006). Revision from translators' point of view. An interview study. *Target*, 18(2), 295–312. doi:10.1075/target.18.2.05shi
- Shih, C. (2013). Translators' "end-revision" processing patterns and maxims: A think-aloud protocol study. *Arab World English Journal*, 4(3), 31–44.
- Shih, C. Y. (2011). Learning from Writing Reflective Learning Journals in a Theory-based Translation Module. *The Interpreter and Translator Trainer*, 5(2), 309–324. doi:10.1080/13556509.2011.10798823
- Shlesinger, M., & Pöschhacker, F. (2011). Aptitude for interpreting. *Interpreting*, 13(1), 1–4. doi:10.1075/intp.13.1.01int
- Shreve, G. M., Angelone, E., & Lacruz, I. (2014). Efficacy of screen recording in the other-revision of translations: episodic memory and event models. *MonTi: Monografías de traducción e interpretación*, 1, 225–245.
- Shreve, G. M. (2002). Knowing translation: Cognitive and experiential aspects of translation expertise from the perspective of expertise studies. In A. Riccardi (Ed.), *Translation studies: Perspectives on an emerging discipline* (pp. 150–171). New York: Cambridge University Press.
- Shreve, G., Angelone, E., & Lacruz, I. (2014). Efficacy of Screen Recording in the Other-Revision of Translations: Episodic Memory and Event Models. *MONTI*, 1, 225–245. doi:10.6035/MonTI.2014.ne1.7
- Sirén, S., & Hakkarainen, K. (2002). Expertise in translation. *Across Languages and Cultures*, 3(1), 71–82. doi:10.1556/Acr.3.2002.1.5
- Skaaden, H., & Wadensjö, C. (2014). Some considerations on the testing of interpreting skills. In C. Giambruno (Ed.), *Assessing legal interpreter quality through testing and certification: The QUALITAS project* (pp. 17–26). San Vicent del Raspeig, Spain: Publicaciones de la Universidad de Alicante. Speechpool. Retrieved from <http://www.speechpool.net/es/>
- Skaaden, H., & Wadensjö, C. (2014). Some considerations on the testing of interpreting skills. In C. Giambruno (Ed.), *Assessing legal interpreter quality through testing and certification: The QUALITAS project* (pp. 17–26). Sant Vicent del Raspeig, Spain: Publicaciones de la Universidad de Alicante.
- Sluijsmans, D., Dochy, F., & Moerkerke, G. (1999). Creating a learning environment by using self-, peer-, and co-assessment. *Learning Environments Research*, 1(3), 293–319. doi:10.1023/A:1009932704458
- Smith, C. D., Worsfold, K., Davies, L., Fisher, R., & McPhail, R. (2013). Assessment literacy and student learning: The case for explicitly developing students 'assessment literacy'. *Assessment & Evaluation in Higher Education*, 38(1), 44–60. doi:10.1080/02602938.2011.598636
- Sokoli, S. (2005). Temas de investigación en traducción audiovisual: La definición del texto audiovisual. In P. Zabala-beascoa, L. Santamaría, & F. Chaume (Eds.), *La traducción audiovisual: Investigación, enseñanza y profesión* (pp. 177–185). Granada, Spain: Comares.
- Stansfield, C. W., & Hewitt, W. E. (2005). Examining the predictive validity of a screening test for court interpreters. *Language Testing*, 22(4), 438–462. doi:10.1191/0265532205lt304oa
- Starke-Meyerring, D. (2008). *Cross-boundary knowledge making in globally networked learning environments*. Paper presented at SUNY Center for Online Collaborative Learning (COIL), Purchase, NY.

Compilation of References

- Stevenson, R., Knott, A., Oberlander, J., & McDonald, S. (2000). Interpreting pronouns and connectives: Interactions among focusing, thematic roles and coherence relations. *Language and Cognitive Processes*, 15(3), 225–262. doi:10.1080/016909600386048
- Srijbos, J.-W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction*, 20(4), 291–303. doi:10.1016/j.learninstruc.2009.08.008
- Strong, M., & Rudser, S. F. (1985). An assessment instrument for sign language interpreters. *Sign Language Studies*, 49(1), 343–362. doi:10.1353/ls.1985.0008
- Swain, M., Brooks, L., & Tocalli-Beller, A. (2002). Peer-peer dialogue as a means of second language learning. *Annual Review of Applied Linguistics*, 22, 171–185. doi:10.1017/S0267190502000090
- Szarkowska, A. (2013). Auteur description: From the director's creative vision to audio description. *Journal of Visual Impairment & Blindness*, 107(5), 383.
- Tang, G. M., & Tithecott, J. (1999). Peer response in ESL writing. *TESL Canada Journal*, 16(2), 20–38. doi:10.18806/tesl.v16i2.716
- Taylor, C. (2014). Intertextuality. In A. Maszerowska, A. Matamala, & P. Orero (Eds.), *Audio description: New perspectives illustrated* (Vol. 112, pp. 29–40). Amsterdam: John Benjamins.
- Tenner, A. R., & DeToro, I. J. (1992). *Total quality management: Three steps to continuous improvement*. Reading, MA: Addison-Wesley.
- Thelen, M. (2005). Going European: A plea for building a network of simulated translation bureaus. In M. Forstner & H. Lee-Jahnke (Eds.), *CIUTI-Forum Paris 2005. Regards sur les aspects culturels de la communication* (pp. 219-236). Frankfurt a/Main: Peter Lang.
- Thelen, M. (2016a). Quality and objectivity of testing and assessing in translator training: Dilemmas and suggestions. In M. Thelen, G.-W. van Egdom, D. Verbeeck, Ł. Bogucki, & B. Lewandowska-Tomaszczyk (Eds.), *Translation and meaning: New series*, 1 (pp. 79-98). Frankfurt a/Main: Peter Lang.
- Thelen, M. (2016b). Collaborative translation in translator training. *Konin Language Studies KSJ*, 4(3), 253-269. Retrieved June 24, 2017, from <http://www.ksj.pwsz.konin.edu.pl/wp-content/uploads/2017/02/KSJ-43-253-269.pdf>
- Thelen, M. (2016a). Quality and objectivity of testing and assessing in translator training: Dilemmas and suggestions. In M. Thelen, G. van Egdom, D. Verbeeck, & B. Lewandowska-Tomaszczyk (Eds.), *Translation and meaning: New serie*, 1 (pp. 79–98). Frankfurt am Main: Peter Lang. doi:10.3726/978-3-653-05899-4/17
- Thelen, M. (2016b). The practice-oriented translator training curriculum: An example. *Current Trends in Translation Teaching and Learning E*, 3, 163–200.
- Thompson, I. (2009). Scaffolding in the writing center: A microanalysis of an experienced tutor's verbal and nonverbal tutoring strategies. *Written Communication*, 26(4), 417–453. doi:10.1177/0741088309342364
- Timarová, Š., & Salaets, H. (2011). Learning styles, motivation and cognitive flexibility in interpreter training. *Interpreting*, 13(1), 31–52. doi:10.1075/intp.13.1.03tim
- Tirkkonen-Condit, S. (2000). Uncertainty in translation processes. In S. Tirkkonen-Condit & R. Jääskeläinen (Eds.), *Tapping and mapping the processes of translation and interpreting: Outlooks on empirical research* (pp. 123–142). Amsterdam: John Benjamins. doi:10.1075/btl.37.13tir

- Tiselius, E. (2009). Revisiting Carroll's scales. In C. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies* (pp. 95–121). Amsterdam: John Benjamins. doi:10.1075/ata.xiv.07tis
- Tobin, L., & Newkirk, T. (Eds.). (1994). *Taking stock: The writing process movement in the '90s*. Portsmouth, NH: Heinemann.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249–276. doi:10.3102/00346543068003249
- Topping, K. (2009). Peer assessment. *Theory into Practice*, 48(1), 20–27. doi:10.1080/00405840802577569
- Torres-Simón, E., & Pym, A. (2016). The professional backgrounds of translation scholars. Report on a survey. *Target. International Journal of Translation Studies*, 28(1), 110–113. doi:10.1075/target.28.1.05tor
- Toury, G. (1995). *Descriptive Translation Studies and beyond*. Amsterdam: John Benjamins. doi:10.1075/btl.4
- Toury, G. (2012). *Descriptive translation studies and beyond* (Revised edition). Amsterdam: John Benjamins. doi:10.1075/btl.100
- Tsagari, D., & van Deemter, R. (Eds.). (2013). *Assessment issues in language translation and interpreting*. Frankfurt am Main: Peter Lang. doi:10.3726/978-3-653-02510-1
- Tsui, A. B. M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing*, 9(2), 147–170. doi:10.1016/S1060-3743(00)00022-9
- Tudge, J. R. H. (1990). Vygotsky, the zone of proximal development, and peer collaboration: Implications for classroom practice. In L. C. Moll (Ed.), *Vygotsky and education: Instructional implications and applications of sociohistorical psychology* (pp. 155–172). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781139173674.008
- University of Helsinki. (2017, July 13). *Line B, BA Degree in a Discipline not Related to Translation*. Retrieved from <http://blogs.helsinki.fi/translationstudies/files/2014/12/structure-of-degree-programme.pdf>
- Valero-Garcés, C., Vigier, F., & Phelan, M. (2014, October). *Holistic versus analytical assessment of legal translations*. Paper presented at QUALETRA Final Conference, Antwerp.
- Vallerand, R. J., Pelletier, L. G., & Koestner, R. (2008). Reflections on Self-Determination Theory. *Canadian Psychology*, 49(3), 257–262. doi:10.1037/a0012804
- Van de Poel, C., & Segers, W. (Eds.). (2007). *Vertalingen objectief evalueren: Matrices en IJkpunten*. Leuven: Acco.
- van Deemter, R., Maxwell-Hyslop, H., & Townsley, B. (2014). Principles of testing. In C. Giambruno (Ed.), *Assessing legal interpreter quality through testing and certification: The QUALITAS project* (pp. 27–39). Sant Vicent del Raspeig, Spain: Publicaciones de la Universidad de Alicante.
- van der Meer, J., Görög, A., Dzeguze, D., & Koot, D. (2017). *Measuring translation quality: From translation quality evaluation to business intelligence*. De Rijp, The Netherlands: TAUS. Retrieved July 6, 2017, from <http://info.taus.net/quality-management-whitepaper?submissionGuid=3d87690e-380e-45cc-9e44-d2013331530f>
- van Doorslaer, L. (2007). Risking conceptual maps. *Target*, 19(2), 217–233.
- Van Steendam, E., Rijlaarsdam, G. C. W., Van den Bergh, H. H., & Sercu, L. (2014). The mediating effect of instruction on pair composition in L2 revision and writing. *An International Journal of the Learning Sciences*, 42(6), 905–927.
- van Zundert, M., Sluijsmans, D., & Van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, 20(4), 270–279. doi:10.1016/j.learninstruc.2009.08.004

Compilation of References

- Vandepitte, S. (2013, March). *Teaching professional standards and quality assurance*. Paper presented at Optimale WP 6, Brussels.
- Vandepitte, S. (2016). Collaboration in translation practices: A course design for peer feedback in translation training. *Konin Language Studies*, 4(4), 289-304. Retrieved from <http://ksj.pwsz.konin.edu.pl>
- Vandepitte, S., & Gonzales, L. (2017). *The language of definitions: conceptualizing technical communication with the eyes of a translation student*. Paper presented at Interdisciplinary knowledge-making: challenges for LSP research, 21st Conference on Language for Specific Purposes, Bergen, Norway.
- Vandepitte, S., & Lisaité, D. (2016). *Expertise perceptions as a factor for effectiveness of peer feedback in translation training?* 8th EST Congress 'Translation Studies: Moving Boundaries', Aarhus, Denmark.
- Vandepitte, S., Maylath, B., Moustén, B., Isohella, S., & Minacori, P. (2016). Multilateral collaboration between technical communicators and translators: A case study on new technologies and processes. *The Journal of Specialised Translation*, 26, 3-17. Retrieved from http://www.jostrans.org/issue26/art_Vandepitte.pdf
- Vandepitte, S. (2016). Collaboration in translation practices: A course design for peer feedback in translation training. *Konin Language Studies*, 4(4), 289–304.
- Venuti, L. (1995). *The translator's invisibility: A history of translation*. London: Routledge. doi:10.4324/9780203360064
- Vercauteren, G. (2007). Towards a European guideline for audio description. In J. Díaz Cintas, P. Orero, & A. Remael (Eds.), *Media for all: subtitling for the deaf, audio description and sign language* (pp. 139–150). Amsterdam: Rodopi.
- Vermeer, H. J. (1978). Ein Rahmen für eine allgemeine Translationstheorie. *Lebende Sprachen*, 23(1), 99–102.
- Vermeiren, H., Van Gucht, J., & De Bontridder, L. (2009). Standards as critical success factors in assessment: Certifying social interpreters in Flanders, Belgium. In C. V. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting: A call for dialogue between research and practice, American Translators Association Scholarly Monograph Series* (Vol. 14, pp. 297–328). Amsterdam: John Benjamins. doi:10.1075/ata.xiv.14ver
- Vigier Moreno, F. J., & Valero Garcés, C. (2017). Shedding light on the grey area of assessment in legal translation: A case study based on a holistic method. *FITISPos International Journal*, 4, 11-27. Retrieved from <http://dspace.uah.es/dspace/handle/10017/29800>
- Vigier Moreno, F. J., & Valdés Garcés, C. (2017). Shedding light on the grey area of assessment in legal translation: A case study based on a holistic method. *FITISPos International Journal*, 4, 11–27.
- Vinay, J.-P., & Darbelnet, J. (1995). *Comparative stylistics of French and English: A methodology for translation* (J. Sager & M.-J. Hamel Trans.). Amsterdam: John Benjamins. (Original work published 1958)
- Vocaley.es. (n.d.). Retrieved from <http://vocaley.es>
- Waddington, C. (2001). Different methods of evaluating student translation: The question of validity. *Meta: Journal des Traducteur/Meta: Translators' Journal*, 46(2), 312–325.
- Waddington, C. (2001). Should translations be assessed holistically or through error analysis? *Hermes*, 26, 15–38.
- Waddington, C. (2004). Should translations be assessed holistically or through error analysis? *Lebende Sprachen*, 49(1), 28–35. doi:10.1515/LES.2004.28
- Wadensjö, C. (1998). *Interpreting as interaction*. London: Longman.

- Wakabayashi, J. (2003). Think-alouds as a pedagogical tool. In B. J. Baer & G. S. Koby (Eds.), *Beyond the ivory tower: rethinking translation pedagogy* (pp. 61–82). Amsterdam: John Benjamins. doi:10.1075/ata.xii.05wak
- Walczak, A., & Fryer, L. (2017b). Vocal delivery of audio description by genre: measuring users' presence. *Perspectives*, 1-15.
- Walczak, A., & Fryer, L. (2017a). Creative description: The impact of audio description style on presence in visually impaired audiences. *British Journal of Visual Impairment*, 35(1), 6–17. doi:10.1177/0264619616661603
- Wallace, M. (2012). *Predictors of successful performance on U.S. Consortium court interpreter certification exams* (Unpublished doctoral dissertation). Universitat d'Alacant, Alicante, Spain. Retrieved from https://rua.ua.es/dspace/bitstream/10045/28364/1/tesis_melissawallace.pdf
- Wallace, M. (2013). Rethinking bifurcated testing models in the court interpretation certification process. In D. Tsagari & R. van Deemter (Eds.), *Assessment issues in language translation and interpreting* (pp. 67–83). Frankfurt am Main: Peter Lang.
- Wallace, M. (2013). Rethinking bifurcated testing models in the court interpreter certification process. In D. Tsagari & R. van Deemter (Eds.), *Assessment issues in language translation and interpreting* (pp. 67–83). Frankfurt am Main: Peter Lang.
- Wallace, M. (2015). A further call to action: Training as a policy issue in court interpreting. *The Interpreter and Translator Trainer*, 9(2), 173–187. doi:10.1080/1750399X.2015.1051769
- Wallace, M. (2017). Lessons from Finland: Diversity and flexibility in the authorization of court interpreters. In M. Kivilehto, A. Liimatainen, A. Nurmi, L. Salmi, A. Viljanmaa, & M. Wallace (Eds.), *Legal translation and court interpreting: Ethical values, quality, competence training (Forum für Fachsprachenforschung)*. Berlin: Frank & Timme.
- Wang, C. (2015, September). *Quantifying assessment criteria in translator training courses: A step forward for translator trainers and translator users to meet*. Paper presented at New perspectives in assessment in translation training: Bridging the gap between academic and professional assessment, London.
- Wang, J.-H., Jemina, N., Goswell, D., & Carmichael, A. (2015). The design and application of rubrics to assess signed language interpreting performance. *The Interpreter and Translator Trainer*, 9(1), 83–103. doi:10.1080/1750399X.2015.1009261
- Wang, K., & Han, C. (2013). Accomplishment in the multitude of counsellors: Peer feedback in translation training. *The International Journal for Translation and Interpreting*, 5(2), 62–75.
- Wang, K., & Han, C. (2013). Accomplishment in the multitude of counsellors: Peer feedback in translation training. *Translation & Interpreting*, 5(2), 62–75.
- Wang, W. (2014). Students' perceptions of rubric-referenced peer feedback on EFL writing: A longitudinal inquiry. *Assessing Writing*, 19, 80–96. doi:10.1016/j.asw.2013.11.008
- Washbourne, K. (2014). The self-directed learner: Intentionality in translator training and education. *Perspective*, 22(3), 373–387. doi:10.1080/0907676X.2013.827225
- Way, C. (2009). Bringing professional practices into translation classrooms. In I. Kemble (Ed.), *The changing face of translation: Proceedings of the eighth annual Portsmouth translation conference* (pp. 131-142). Portsmouth, UK: University of Portsmouth.

Compilation of References

- Way, C. (2016). Intra-university projects as a solution to the simulated/authentic dilemma. In D. Kiraly, L. R uth, & M. Wiedmann (Eds.), *Towards authentic experiential learning in translator education* (pp. 147-159). G ttingen: V&R unipress/Mainz University Press.
- Way, C. (2008). Systematic assessment of translator competence: in search of Achilles' heel. In J. Kearns (Ed.), *Translator and interpreter training: Issues, methods and debates* (pp. 88–103). London: Continuum International.
- Way, C. (2008). Systematic assessment of translator competence: In search of Achilles' heel. In J. Kearns (Ed.), *Translator and interpreter training: Issues, methods and debates* (pp. 88–103). London: Continuum.
- Way, C. (2014). Structuring a legal translation course: A framework for decision making in legal translator training. In L. Cheng, K. K. Sin, & A. Wagner (Eds.), *The Ashgate handbook of legal translation* (pp. 135–154). London: Routledge.
- Way, C. (2014). Structuring a legal translation course: A framework for decision-making in legal translator training. In L. Cheng, K. K. Sin, & A. Wagner (Eds.), *The Ashgate handbook of legal translation* (pp. 135–152). Farnham, UK: Ashgate.
- Weidmann, A. (2014). Critiquez-nous! Mais critiquez-nous bien. Pour une critique grand public des traductions/adaptations audiovisuelles. *L' cran traduit*, 3, 45-71.
- Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer-Verlag. doi:10.1007/978-1-4612-4948-1
- Weiss, T. (1999). The implications of translation for professional communication. In C. R. Lovitt & D. Goswami (Eds.), *Exploring the rhetoric of international professional communication: An agenda for teachers and researchers* (pp. 277–292). New York: Baywood. doi:10.2190/ETRC12
- Wells, G. (1999). *Dialogic inquiry: Toward a sociocultural practice and theory of education*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511605895
- Wiersma, W., & Jurs, S. G. (1990). *Educational measurement and testing*. Boston: Allyn & Bacon.
- William, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Mahwah, NJ: Erlbaum.
- Williams, M. (2009). Translation quality assessment. *Mutatis Mutandis*, 2(1), 3–23.
- Williams, M. (2009). Translation quality assurance. *Mutatis Mutandis*, 2(1), 3–23.
- Williams, M., & Burden, R. L. (1997). *Psychology for language teachers: A social constructivist approach*. Cambridge, UK: Cambridge University Press.
- Wilss, W. (1976). Perspectives and limitations of a didactic framework for the teaching of translation. In R. W. Brislin (Ed.), *Translation: Applications and research* (pp. 117–137). New York: Gardner Press.
- Wilss, W. (1996). *Knowledge and skills in translator behavior*. Amsterdam: John Benjamins. doi:10.1075/btl.15
- Wissmath, B., Weibel, D., & Groner, R. (2009). Dubbing or subtitling? Effects on spatial presence, transportation, flow, and enjoyment. *Journal of Media Psychology*, 21(3), 114–125. doi:10.1027/1864-1105.21.3.114
- Witter-Merithew, A., Taylor, M., & Johnson, L. (2001). Guided self-assessment and professional development planning: A model applied to interpreters in educational settings. In C. Nettles (Ed.), *Proceedings of the 17th national conference of the registry of interpreters for the deaf* (pp. 153-226). Alexandria, VA: RID Publications.

- Witter-Merithew, A., Taylor, M., & Johnson, L. (2001). Guided self-assessment and professional development planning: A model applied to interpreters in educational settings. In C. Nettles (Ed.), *Proceedings of the 17th National Conference of the Registry of Interpreters for the Deaf* (pp. 153-226). Alexandria, VA: RID Publications.
- Wu, S. C. (2010). *Assessing simultaneous interpreting: A study on test reliability and examiners' assessment behavior* (Unpublished doctoral dissertation). Newcastle University, UK.
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., & Norouzi, M. (2016). *Google's neural machine translation system: Bridging the gap between human and machine translation*. Retrieved June 9, 2017, from <https://arxiv.org/pdf/1609.08144.pdf>
- Wu, F. S. (2013). How do we assess students in the interpreting examination? In D. Tsagari & R. van Deemter (Eds.), *Assessment issues in language translation and interpreting* (pp. 15–33). Frankfurt am Main: Peter Lang.
- Yang, M., Badger, R., & Yu, Z. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing, 15*(3), 179–200. doi:10.1016/j.jslw.2006.09.004
- Yeh, S., & Liu, M. (2006). A more objective approach to interpretation evaluation: Exploring the use of scoring rubrics. *Journal of the National Institute for Compilations and Translations, 34*(4), 1–50.
- Yeh, S.-P., & Liu, M. (2006). 口譯評分客觀化初探:採用量表的可能性 [A more objective approach to interpretation evaluation: Exploring the use of scoring rubrics]. *Journal of the National Institute for Compilation and Translation, 34*(4), 57–78.
- Zhang, S. (1995). Reexamining the affective advantage of peer feedback in the ESL writing class. *Journal of Second Language Writing, 4*(3), 209–222. doi:10.1016/1060-3743(95)90010-1
- Zhong, Y. (2005). Plan-based translation assessment: An alternative to the standard-based cut-the-feet-to-fit-the-shoes style of assessment. *Meta, 50*(4).
- Zili, D. (2009). An empirical study on campus internet-based peer review and interactive translation. *Chinese Translators Journal, 3*, 44–49.
- Zimmerman, B., & Schunk, D. (2011). Self-Regulated Learning and Performance: An Introduction and an Overview. In B. Zimmerman & D. Schunk (Eds.), *Handbook of Self-Regulation of Learning and Performance* (pp. 1–12). New York: Taylor & Francis.
- Zwischenberger, C., & Pöchhacker, F. (2010). Survey on quality and role: Conference interpreters' expectations and self-perceptions. *Communicate (High Wycombe), 53*. Retrieved from <http://www.aiic.net/ViewPage.cfm/article2510.htm>

About the Contributors

Elsa Huertas Barros is a Lecturer in Translation Studies in the Department of Modern Languages and Cultures at the University of Westminster. Elsa holds a PhD in Translation from the University of Granada. Elsa's main research interest include translator education, translator competence, assessment practices, collaborative learning and student-centred approaches. Elsa has published her work in international journals including *The Interpreter and Translator Trainer (ITT)*, *The Journal of Specialised Translation (JoSTrans)*, and the *International Journal of Legal Discourse (IJLD)*. Elsa has also published book chapters in edited volumes such as *Translation and Meaning*, by Peter Lang, *Phraseology in Legal and Institutional Settings: A Corpus-based Interdisciplinary Perspective*, by Routledge, and *Employability for Languages: a Handbook*. Elsa has recently co-edited a special issue of the *ITT* (12,1) on *New Perspectives in Assessment in Translator Training*.

Sonia Vandepitte is a full professor at the Department of Translation, Interpreting and Communication at Ghent University and head of its English section. She teaches English, translation studies, and translation into and from Dutch. She also coaches student translation companies. Publication topics include causal expressions in language and translation, methodology in translation studies, translation competences, anticipation in interpreting, international translation training projects and translation and post-editing processes. She is currently involved in eye-tracking research into reading and translation processes of translation problem-solving, and she is also investigating peer feedback and other collaborative forms of learning in translation training.

* * *

Erik Angelone is Associate Professor of Translation Studies at Kent State University. He received his Ph.D. in Translation Studies from the University of Heidelberg. His current research interests include process-oriented translator training, cognitive processes in translation, and intercultural communicative competence.

Henri Bloemen is an Associate Professor teaching Translation Studies, Translation Theory, Literary Translation and Law Translation at KULeuven, Belgium. Publications in the fields of Theory of Translation, Ethics of Translation, Didactics of Translations. He was one of the promoters of the PETRA-E Project on the competence of literary translators.

Ya-Yun Chen (PhD) teaches on the post-graduate translating and interpreting programme at Newcastle University. Her research interests include interpreter and translator pedagogy, interpreting studies from a social-psychological and affective perspective, and audio-visual and game translation. Her main publications are in the areas of Chinese-English translation and interpreting/translation pedagogy. She is an award-winning translator and has translated more than 50 English books into Chinese. She also has extensive experience in the localization industry.

Si Cheng has recently finished her PhD research at the Centre for Translation and Intercultural Studies at the University of Manchester. Her research focuses on translation competence and translation pedagogy.

José Tomás Conde Ruano teaches Translation and Interpreting at the University of the Basque Country (UPV/EHU) and works within the framework of several research groups, such as TRALIMA, PETRA and GENTT. PhD in Translation and Interpreting from the University of Granada, his thesis deals with translation quality and evaluation and was the runner-up in the 2009-2011 Prize for the Doctoral Thesis, granted by the Iberian Association of Translation and Interpreting Studies (AIETI). Some of his last works have appeared on *Parallèles*, *Cadernos de Tradução*, *Trans*, *Hermeneus*, *Ibérica*, *Across Languages and Cultures*, *Meta* and *JosTrans*, inter alia. Today, he is interested in shedding some light on the processes involved in the assessment of audiovisual translations.

Louise Fryer is one of the UK's most experienced describers. She was the BBC's describer for the Audetel project, which piloted TV AD in the 1990s and she helped to establish one of the first description units for screen AD in the UK. She has been a member of the Royal National Theatre's audio description team since it was set up in 1993. She has served on the board of the UK's Audio Description Association (ADA) for whom she is an accredited trainer and has trained describers for stage and screen in the UK and elsewhere. She was the access advisor and described the BAFTA-nominated film *Notes on Blindness* (Middleton/Spinney, 2017). She is a teaching fellow in AVT at University College London (UCL). She has published extensively on AD in academic journals and is the author of *An Introduction to Audio Description: A Practical Guide*.

Chao Han is associate professor in the College of International Studies, Southwest University in China. His research interests include testing and assessment of language interpreting, evidence-based interpreting studies, and research methodology.

Silvia Hansen-Schirra, Dipl.-Übers., Dr. phil., PD, is a full professor of English linguistics and translation studies at Johannes Gutenberg University Mainz in Gernersheim, Germany. Her main research interests include specialized communication, text comprehensibility, post-editing, translation process and competence research. As fellow of the Gutenberg Research College she is the director of the Translation & Cognition (TRACO) Center in Gernersheim and co-editor of the online book series "Translation and Multilingual Natural Language Processing".

Joleen Hanson, Ph.D., is Associate Professor of English at the University of Wisconsin-Stout (Menomonie, WI, USA) where she teaches courses in linguistics, transnational professional communication, rhetoric and composition, and second-language writing. Her research focuses on writing pedagogy, the

About the Contributors

design of globally networked learning environments, and the development of language awareness and intercultural competence among professional communicators.

Carmen Heine, PhD, Associate Professor, is working with English Business Communication at the Department of English, School of Communication and Culture at Aarhus University. Her research areas are Writing Research and Translation Studies with a particular focus on cognitive aspects of text production process research.

Sascha Hofmann, Dipl.-Übers., Dr. phil., is Associate Professor for English Linguistics and Translation Studies at the FTSK at the Johannes Gutenberg University of Mainz. His main research interests include general applied translation studies, translation process research, as well as all aspects of translator education. As member of the of the Translation & Cognition (TRACO) Center he is together with Don Kiraly principal researcher in the field of translation process related competence development.

Hendrik J. Kockaert lectures French Linguistics, Terminology, Legal Translation, and Translation Technology at KU Leuven, Faculty of Arts, in Antwerp. Since August 2015, he is Dean of the Faculty of Arts on Campus Sint Andries in Antwerp. He is a Research Associate at the University of The Free State, Republic of South-Africa. He is a certified LICS Auditor for granting ISO 17100 certification to translation services, and he is the Editor-in-Chief of The Journal of Internationalisation and Localisation [JIAL], published by John Benjamins Publishing Company. He is the Chairperson of ISO TC 37 SC 1 and a member of NBN, the Belgian Standardization Institute. He is an expert in the development of ISO terminology standards. He is a certified ECQA (European Certification and Qualification Association) Terminology Manager Trainer and Job Role Committee Provider. He was the coordinator of the following projects financed by the European Commission: QUALETRA (JUST/2011/JPEN/AG/2975), and LIT Search (JUST/2013/JPEN/AG/4556). He publishes in the areas of terminology, translation quality assurance, and translation technology. Together with Prof. Dr. Winibert Segers and Televic, he develops translationQ, a CAT tool that delivers an automated evaluation and revision process.

Gary Massey is Director of the IUED Institute of Translation and Interpreting at ZHAW Zurich University of Applied Sciences. A principal investigator and co-investigator on federally funded research projects on translation teaching, cognitive translation processes and workplace ergonomics, his major publications, research and teaching interests include translator competence development, process-oriented and collaborative translator education and translation quality assessment.

Mari Pakkala-Weckström is a Senior Lecturer of English Translation at the Department of Modern Languages, University of Helsinki.

Jasper Pauwels received his BA in Translation (Vertaalacademie) from the Zuyd University of Applied Sciences in 2015 and received his MA in translation from the University of Antwerp in 2017. His master's thesis research investigated and compared the translation processes of third-year students from the BA Applied Linguistics from the University of Antwerp and of third-year students from the BA in translation from the Zuyd University of Applied Sciences. Upon graduation from the University of Antwerp, he started to pursue a career as a free legal translator.

Moritz Schaeffer received his PhD from the University of Leicester in translation studies and has since worked as a research assistant at the Center of Research and Innovation in Translation and Translation Technology (CRITT) (Copenhagen Business School), the Institute for Language, Cognition and Computation (University of Edinburgh) and the National Institute of Informatics (Tokyo). He is currently a senior research associate at the Faculty of Translation Studies, Linguistics and Cultural Studies at the Johannes Gutenberg University of Mainz.

Iris Schrijver is assistant professor (tenure-track) in Translation Studies and Translation Spanish into Dutch at the department of Applied Linguistics of the University of Antwerp. She has been awarded the EST Young Scholar Prize 2016 for her PhD research into the effects of writing training on translation performance. Her research focuses on translation competence acquisition and cognitive aspects of text-production processes such as writing and translation.

Winibert Segers is affiliated to KU Leuven. He lectures in French and Dutch on administrative and medical texts. He completed a doctorate on the untranslatability of writings by the French philosopher, Jacques Derrida. He publishes in the following research areas: translation evaluation, translation didactics, translation ethics, translation philosophy, translation theory and translations of travelogues.

Anke Tardel studied at the Faculty of Translation Studies, Linguistics and Cultural Studies in Germersheim, where she received her B.A. and her M.A. degree in Translation. Besides her studies, she has been a student research assistant since 2012, supporting various eye tracking studies and research projects at the Translation & Cognition (TRACO) Center. As of October 2017, she is a PhD student and holds a research and teaching position at the same faculty. Her research interests include translation process research, translation revision, post-editing machine translation, translation technologies, and cognitive translation studies.

Marcel Thelen, independent researcher, and senior lecturer in terminology and translation (ret.), was affiliated to the Maastricht School of Translation and Interpreting of Zuyd University of Applied Sciences for 33 years. He was Head of School from 2007 to 2012. From 2004 to 2006 he was chairman of the NEN-commission for translation services for the European standard EN 15038; from 2012 to 2014 he took part in the development of ISO 17100. He holds a PhD from Vrije Universiteit Amsterdam on The Structure of the Lexicon. Incorporating a cognitive approach in the TCM lexicon, with applications to lexicography, terminology and translation. He is co-founder and co-organiser of the International Maastricht-Łódź Duo Colloquium on “Translation and Meaning”, and vice-chairman of NL-Term, the Dutch-Flemish association for Dutch terminology. He publishes and presents widely on translation, terminology, lexicology, quality management, translation didactics and curriculum building.

Gys-Walt van Egdom holds a doctor’s degree in Linguistics and Literary Studies and a master’s degree in Translation Studies. He is a lecturer and researcher at Zuyd University of Applied Sciences (Maastricht School of Translation and Interpreting) and is also affiliated to Vrije Universiteit Brussel. He is researcher at the Centre of Expertise Translation Didactics and Centre of Expertise International Relationship Management. His research interests include translation didactics, translation evaluation, translation ethics, machine translation and post-editing, multilingual sentiment analysis and language philosophy.

About the Contributors

Heidi Verplaetse currently teaches English in the domains of translation (medical/scientific, business and journalistic texts), writing (business and journalistic texts) and external communication at KU Leuven, Campus Sint-Andries in Antwerp. Her earlier academic teaching included English phonology & phonetics and pronunciation at Ghent University, where she subsequently also worked on research projects in the domain of knowledge management based on automatic retrieval of (causal and related) semantic relations and anaphora. Her PhD in English linguistics was situated in the domain of modality as expressed in informal spoken data from the British National Corpus. She subsequently also published in the domain of business communication. Her current research interests are situated in the domains of specialized (medical) translation, translation quality and the use of CAT tools, as well as trans-editing of journalistic texts and (the evolution of) translation work and processes in students.

Juliet Vine is a Senior Lecturer in Translation and Interpreting at the University of Westminster's Department of Modern Languages and Cultures. Juliet's research interests are pedagogical research and contrastive rhetoric focusing on Chinese and western rhetorical traditions. Juliet has presented her work at international conferences such as didTRAD and the first World Interpreter and Translator Training Association (WITTA) Congress. Juliet has published book chapters in edited volumes such as *Translation and Meaning*, published by Peter Lang and *Employability for Languages: a Handbook*. Juliet has recently co-edited a special issue of the *Interpreter and Translator Trainer* (ITT, 12,1) on *New Perspectives in Assessment in Translator Training*.

Melissa Wallace received her Ph.D. in translation and interpreting studies from the Universidad de Alicante, Spain. A certified court interpreter and certified healthcare interpreter, Wallace is currently serving a 5-year term on the Licensed Court Interpreter Advisory Board of the Judicial Branch Certification Commission for the Supreme Court of Texas. She is an active member of the Home for Trainers Webinars Work Group at the National Council on Interpreting in Health Care, and is a board member of the Society for the Study of Translation and Interpretation (SSTI), the non-profit educational and research foundation of the National Association of Judicial Interpreters and Translators. Her research focuses on indicators of aptitude on court interpreter certification exams, accreditation exam models, and interpreter and translator training. Wallace carried out research on court interpreting certification models in Finland in 2016 as the Fulbright-University of Tampere Scholar. She is an Assistant Professor of TI Studies at the University of Texas at San Antonio where she directs the graduate certificate program in translation and interpreting studies.

Regine Wieder is Senior Lecturer in Organisational Communication in the IAM Institute of Applied Media Studies and Director of the MA Specialisation in Organisational Communication at ZHAW Zurich University of Applied Sciences. Her principal publications, research and teaching interests cover various aspects of corporate communications and strategic communication management, with a particular emphasis on their international dimensions.

Bert Wylin (MA Applied Linguistics – Roman languages) has both an academic and a business profile. He works at the K.U.Leuven since 1993, doing research in the fields of technology supported teaching and (language) learning. In 2001, he founded a spin-off company, now merged into Televic Education, developing and servicing educational e-learning and e-assessment projects worldwide and specializing in languages and computer assisted language learning (CALL).

Index

A

Accent 118-119, 161
 Affordances 29, 164
 Attitude 31, 92, 277, 283, 290, 328

C

Characteristics 2-3, 20, 24, 27-28, 32, 54, 90-92, 95,
 101, 120, 130, 159, 341
 Coaching 62, 64
 Creativity 59, 76, 120, 147, 158
 Cultural Diversity 323

D

Diversity 143, 316, 323, 352-353

E

Exploring 1, 57, 61, 76, 199, 315

I

Inclusion 162-163, 187, 189-190, 246, 258

M

Mentoring 261, 265
 Methodology 134, 136-137, 250, 292, 341
 Motivation 58, 63, 74, 106, 120, 271, 296, 306-308,
 310-311
 Multilingual 58, 61, 63, 75-77, 191, 279

P

Peers 91, 93, 99, 102-103, 105, 107, 305, 308, 316,
 319, 321, 337-340, 342-343, 347-351
 Planning 61, 221, 249, 273, 339, 353
 Policies 118, 135
 Professional Development 62, 64, 265, 272, 288

Q

Quality Assurance 1-2, 5-6, 8-10, 12, 14, 20, 24, 57,
 62, 74-76, 78, 113, 156, 180, 183, 240, 316, 353

R

Reasoning 255

S

Scaffolding 64, 330, 337-339, 341, 343, 351-353
 Storytelling 60, 75

T

Teacher Training 246
 Teaching 28, 30, 38, 54, 94, 99, 116, 129, 157, 161, 200,
 212-213, 215, 246-247, 249-251, 277, 289-290,
 318, 330, 338-339, 341, 344, 349, 352
 typologies 32-33