Learning Language through Task Repetition

Edited by
Martin Bygate

Task-Based Language Teaching

11

John Benjamins Publishing Company

Learning Language through Task Repetition

Task-Based Language Teaching: Issues, Research and Practice (TBLT)

ISSN 1877-346X

Task-Based Language Teaching (TBLT) is an educational framework for the theory and practice of teaching second or foreign languages. The TBLT book series is devoted to the dissemination of TBLT issues and practices, and to fostering improved understanding and communication across the various clines of TBLT work.

For an overview of all books published in this series, please see http://benjamins.com/catalog/tblt

Editors

Martin Bygate
Lancaster University

John M. Norris
Educational Testing Service

Kris Van den Branden
KU Leuven

Volume 11

Learning Language through Task Repetition
Edited by Martin Bygate

Learning Language through Task Repetition

Edited by

Martin Bygate
University of Lancaster

John Benjamins Publishing Company
Amsterdam / Philadelphia

DOI 10.1075/tblt.11

Cataloging-in-Publication Data available from Library of Congress.

ISBN 978 90 272 0113 3 (HB)
ISBN 978 90 272 0114 0 (PB)
ISBN 978 90 272 6378 0 (E-BOOK)

John Benjamins Publishing Company · https://benjamins.com

Table of contents

Series Editors' Preface vii

Introduction 1
Martin Bygate

CHAPTER 1
Task repetition for language learning: A perspective from skill acquisition theory 27
Robert DeKeyser

CHAPTER 2
Grammatical structures and oral fluency in immediate task repetition: Trigrams across repeated performances 43
Nel de Jong and Philip Tillman

CHAPTER 3
The effects of task repetition and task complexity on L2 lexicon use 75
YouJin Kim, Scott Crossley, YeonJoo Jung, Kristopher Kyle and Sanghee Kang

CHAPTER 4
Discourse performance in L2 task repetition 97
Zhan Wang and Gaowei Chen

CHAPTER 5
The impact of intra-cultural and inter-cultural task repetition on interaction 117
Scott Aubrey

CHAPTER 6
Effects of task type, task-type repetition, and performance criteria on L2 oral production 143
Xingchao Hu

CHAPTER 7
The effects of awareness-raising through stimulated recall on the repeated performance of the same task and on a new task of the same type 171
Chris Sheppard and Rod Ellis

CHAPTER 8
Perform, reflect, recycle: Enhancing task repetition in second language speaking classes 193
Tony Lynch

CHAPTER 9
Second language learning through repeated engagement in a poster presentation task 223
Emi Kobayashi and Masaki Kobayashi

CHAPTER 10
Mediating input-based tasks for beginner learners through task repetition: A sociocultural perspective 255
Natsuko Shintani

CHAPTER 11
Understanding benefits of repetition from a complex dynamic systems perspective: The case of a writing task 279
Ryo Nitta and Kyoko Baba

CHAPTER 12
Task repetition or task iteration? It does make a difference 311
Diane Larsen-Freeman

External reviewers 331

Subject index 333

Series Editors' Preface

There is a lot of intuitive appeal to the idea that task repetition enhances development. When human beings try to acquire a complex skill (like riding a bike, driving a car, swimming, or delivering a keynote), their very first attempt is highly unlikely to be their best. To the contrary, when a complex task is repeated a number of times, chances are considerable that people gradually get better, especially when they are provided with interactional support and feedback, and/or reflect upon their own performance.

Paradoxically, this line of reasoning implies that there may be no such thing as exact task repetition. Even if students are set what is seemingly the very same task, their performance will not be identical to their previous one. As Martin Bygate makes clear in his excellent introduction to this volume, "task repetition" can refer to a lot of things, but it usually does not refer to the exact repetition of a particular performance of exactly the same task. For one, the task and its accompanying instructions are seldom identical across subsequent performances, even in studies that focus on the effects of task repetition. What needs to be repeated is usually some aspect of a particular task. For instance, after a first attempt at a directions-giving task, learners may be asked to give directions about the same route to another interlocutor, to perform the same task type but with a different destination, to perform the same task under more stressful conditions (e.g., under time pressure), and the like. Secondly, even if tasks (or parts of tasks) may be the same on paper, and so involve very similar instructions, from the moment people engage with tasks, variation is bound to occur. People, unlike machines, do not just copy and paste their communicative performances. In fact, for learning to happen, this is crucial: Learning, ultimately, is the result of variation and change.

Cognition-based research into learning processes provides substantial evidence for practice and repetition effects (cf. Hattie & Yates, 2014; Weinstein, Madan, & Sumeracki, 2018). For instance, there is robust empirical evidence for the impact of deliberate practice, distributed practice, and interleaving, all of which involve significant amounts of task repetition. Both storage strength (the extent to which a memory is truly embedded in the mind) and retrieval strength (the ease with which a memory can be retrieved) appear to be positively affected by task repetition in multiple ways. Interestingly, much of this research suggests that most authentic, real-life tasks

may be composed of multiple minitasks and involve a considerable host of subskills. Repetition, then, may allow students to automatize certain of these minitasks or subskills, freeing their mind to focus on another minitask or subskill during subsequent performances. Or task repetition may promote the fluent and effective integration of several of these minitasks into a more holistic performance. This is a fascinating thought that is taken up in several chapters of this volume, both from a theoretical and empirical perspective. It may actually shake up the whole concept of a task, which in many seminal articles on TBLT, tends to be treated as a one-dimensional unit, rather than a composite concept.

Equally fascinating is the thought that even if in subsequent performances some features of a task may be relatively stable, the interaction to which the task gives rise will strongly differ across performances. Particular types of interactional variations (such as the degree of scaffolding or modelling offered by a teacher, the feedback provided by the interlocutor, the types of questions that the L2 speaker is asked, and the specific output that the L2 speaker will produce during subsequent performances) may have a direct or indirect impact on the learning that results from the repeated task performance. The studies reported in this volume shed light on the complexity of the interplay between task repetition and the concomitant modification of output production and interactional support. Descriptions of task sequences, such as they tend to be produced in curricula or syllabi, may touch upon only a fragment of the many variables that are involved in the kind of sequencing of task performance that gives rise to language learning.

As much as there is to be learnt from the theoretical and empirical chapters in this volume with regard to handling task repetition in task-based language teaching, the volume also shows that we may have only just begun unravelling the wonders of task repetition. One crucial question is how we measure the progress students make across task repetitions, and what, ultimately, is the relationship between variation in performance and sustainable learning. If some learners are found to have produced slightly more complex sentences (from a syntactic point of view) in task performance 2, or have been found to produce a significantly slighter amount of hesitations (adding to their fluency scores), what exactly can these learners be said to have learnt? Would it require additional task performances to measure the true impact of task repetition on learning, and if so, would that subsequent task performance not constitute another hotspot for further learning? In addition, how much of what is truly learnt can be discerned using the methods and instruments that are typically employed in task repetition research? Do we need more sophisticated tools or more varied perspectives to document learners' progress in task-related behavior and the mental processing involved?

Finally, for all its richness, this volume contains chapters that approach task repetition from the perspective of one particular target language. However, most of the

tasks that second language learners are asked to perform are familiar to them in the sense that they have produced them (and may have performed them many times) in their L1 (or any other language that they use frequently). Both in the theory-making and the empirical research into task repetition, cross-linguistic influences may constitute another untapped, but potentially powerful, area of research.

As concluded by its editor, this volume on task repetition is overdue, and its content is clearly provocative for researchers and practitioners interested in the many nuances of this deceptively simple idea. There clearly is a need for much more research into this promising area, and this volume provides an excellent foundation for inspiring future work.

Kris Van den Branden
John M. Norris

References

Hattie, J., & Yates, G. (2014). *Visible learning and the science of how we learn*. New York: Routledge.

Weinstein, Y., Madan, C., & Sumeracki, M. (2018). Teaching the science of learning. *Cognitive Research: Principles and Implications*, 3. https://doi.org/10.1186/s41235-017-0087-y

Andy Warhol

'I started repeating the same image because I like the way the repetition changed the same image.' (from 'A conversation with Andy Warhol' by Gerard Malanga, in K.Goldsmith (ed.) 1971. I'll be your mirror: the selected Andy Warhol interviews. London: Hachette UK)

Ricky O'Bannon

'If you listen to music, you instinctively know that a song sounds different the tenth time you hear it from the first. Repetition is an often overlooked yet powerful part of the way we process music, whether that music is a classical symphony or that Taylor Swift song we just can't seem to get out of our head.'

https://www.bsomusic.org/stories/the-power-of-musical-repetition/

Margulis, E.M. (Director, Music Lab, University of Arkansas)

'Research has also shown that listeners shift their attention across musical repetitions, focusing on different aspects of the sound on each new listen. You might notice the melody of a phrase the first time, but when it is repeated your attention shifts to how a guitarist bends a pitch. 'Why we love repetitions in music.' TEDEd Lessons worth sharing.

Graham Fitch

'The trick with repetitions [in learning a piece of music] is to focus the mind on something very specific, varying the focus with each repetition. If you are refining a tricky spot in a piece you have already learned, your brain should be consuming a lot of energy as you concentrate on what is necessary to edge the passage closer to where it needs to be, rather than simply hacking away at it hoping it will eventually yield (this requires far less concentration).

http://www.practisingthepiano.com/how-to-manage-repetition-in-practice/

Pablo Casals

'It is not a mechanical routine but something essential to my daily life. I go to the piano, and I play two preludes and fugues of Bach. I cannot think of doing otherwise. It is a sort of benediction on the house. But that is not its only meaning for me. It is a rediscovery of the world of which I have the joy of being a part. It fills me with awareness of the wonder of life, with a feeling of the incredible marvel of being a human being. The music is never the same for me, never. Each day is something new, fantastic, unbelievable. That is Bach, like nature, a miracle!'

Introduction

Martin Bygate
University of Lancaster

Before returning home at the end of her year of study at a UK university, an Italian student remembered that she needed to claim back the deposit she had paid for her room in a university residence. She went to the bursar's office of the residence, and negotiated the quaint university term 'battels' to refer to the deposit. She also negotiated the information that she had to go to a central university office. So she headed for the University accommodation office, where she succeeded in using the word 'battels', and also in getting re-directed to the University finance office. There, and without further struggle, she happily got her refund. The three encounters were a real-life illustration of the concept of 'task repetition'.

It is unlikely that the three encounters were all linguistically identical: apart from using the word 'battels', she was unlikely to express herself in exactly the same way each time, and the staff behind the counters would not all have produced the same utterances. However, much of the information she had to give each time would have been the same, some of the vocabulary would have been the same, and on each occasion she would have used utterances with similar functions, such as explaining, requesting, and providing her identity. The interest in the sequence of the three encounters is that, apart from getting her money back, the student also had the opportunity to maintain and improve aspects of her English. Task repetition occurs in real life, and could be a useful site for language learning.

This anecdote was reported to me by one of my students, and it aroused my curiosity about task repetition as a naturally occurring context for language learning, and hence as a potentially interesting research area. In this account, the term 'task repetition' does not refer to the precise repetition of the language used. As in the example, tasks are by definition 'unscripted', which means that when a task is repeated, participants choose their own language to express their meanings, and this language can vary. Indeed, the sequence of utterance types can also vary. Rather what is repeated is a given configuration of purposes and a set of content information. In these terms, task repetition often occurs in real-life, for instance whenever we recount an

DOI 10.1075/tblt.11.intro

experience, describe a familiar route or location, repeat the same service transaction, exchange small talk about the same weather patterns, or discuss the same problems – health, social, political, or other.

In pedagogical terms, the definition provided above allows for a number of different types of task repetition. Since tasks can be jointly carried out by teachers with groups or with the whole class, task repetition might involve a teacher repeating a task with a class, or indeed a task being carried out firstly by groups and then subsequently in whole class format. The definition allows for the possibility of task repetition occurring when the task retains the same purpose and content but with the content reorganised in some way (for instance two or more picture differences tasks all with the same features in the pictures but differently configured). And the definition also covers tasks where material is iterated first orally and then in writing. I am clarifying my definition here because task repetition has been interpreted in different ways. In this chapter I will first consider the construct of 'task repetition' and situate task repetition research within TBLT. Then after briefly reviewing early research, I will consider the potential significance of the concept for language teaching, and conclude with an overview of this volume.

The construct of 'task repetition'

It has been remarked (see for instance Larsen Freeman this volume) that, strictly speaking, nothing can be precisely repeated, including tasks. Indeed this is true of task repetition as defined above. The definition, 'repetition of a given configuration of purposes, and a set of content information' allows for variation from one iteration to another. On the one hand the definition simply requires a minimum recurring set of purposes and content information. Provided that a minimum recurs, the definition allows for possible substitutions of language or moves, as well as additions or omissions. For instance, the student requesting a refund of her battels might have varied her vocabulary from one iteration to the next; she might have provided different amounts of background information; she might have greeted and bade farewell in different ways; if in one of the encounters she knew one or more of her interlocutors personally then she would likely have added some social talk; she would certainly have thanked differently at the end of the third encounter. Furthermore, it is quite likely that she herself was differently primed for each encounter. For instance, as she went from one office to the next, the key sequence of utterances introducing, backgrounding, and then presenting her request would likely have become clearer and clearer in her mind, and increasingly more language might have been accessible, such as relevant vocabulary, and potential requesting formulae. Nevertheless, despite these potential differences, the concept of task repetition implies the possibility of our agreeing that key elements

of purpose and content material recur. The interest in the topic is precisely what kinds of changes occur when learners engage in task repetition, and how they might relate to language development and that is the focus of this volume.

'Task' is a rich and potentially powerful construct, central to the development of task-based language teaching. Perhaps its most important role is to provide a meaningful context for language use and hence for learning that can be used by teachers. The complex of meanings activated in order to breathe life into tasks and then carry them out in order to achieve one or more purposes is a powerful basis for the use of inferencing skills, which are important in enabling learners to understand new language and to make themselves understood using language they have difficulty in marshalling. Furthermore, tasks can help learners to shape their functional grasp of language: language is given purpose by the need to ensure communication with interlocutors, and with that purpose, is shaped into discourse patterns. Thus tasks can be seen as constituting a kind of shell that can provide a functional niche for each language feature that goes to make up the learners' discourse, and thus help learners relate the features to each other. So for example, tasks involving the telling of stories are likely to bring into contextually defined functional relationships the simple past tense, anaphoric reference of articles and pronouns, adverbials of time, place, and purpose, and conjuncts helping to mark sequentiality. In this respect, they can be seen as having a similar status to the interactional routines which Bruner (1983) interpreted as being highly significant for child language acquisition, and which he referred to as 'formats'.

Tasks as attentional frames

Bruner identified these formats during many hours of observation of parent-child dyads. They took the form of peekaboo games, meal time interactions, and interactional routines that occurred at points in the day such as bath time, bed time, and before going out in the stroller. As Tomasello (2003) remarks:

> Social interactional routines such as feeding, diaper changing, bathing, interactive games, book reading, car trips, and a host of other activities constitute the formats – joint attentional frames – within which children acquire their earliest linguistic symbols […]. (p. 88)

These routines provide important support for learning to take place:

> Within joint attentional frames, adults often make linguistic utterances in an attempt to exhort children to attend to certain aspects of the shared situation. In attempting to comply with these exhortations – that is, in attempting to comprehend the adult's communicative intention as expressed in the utterance – children use all kinds of

> interpretive strategies based on the pragmatic assumption that utterances are somehow *relevant* to the ongoing interaction […]. In the social-pragmatic view, word learning does not consist in the child engaging in a reflective cognitive task in an attempt to make correct mappings of word to world, but rather it emerges naturally from situations in which children are engaged in social interactions in which they are attempting to understand and interpret adult communicative intentions as expressed in utterances. (p. 89)

Thus the shared attentional frame provides the key reference point for communicative success, and through communication, learning. Tasks can be seen as fulfilling broadly the same function for second language acquisition: language learning tasks, too, offer shared attentional frames – tasks such as spot-the-difference tasks, story-telling tasks, prioritising tasks, description tasks, tasks involving directions, recommendation tasks, decision-making tasks, and so on. So too do tasks involving teacher-class interaction, such as completing a weather chart, following teacher directions on maps, filling in plans of rooms, or labelling zoos from the teacher's description.

However, the contribution of the 'shared attentional frames' to child language acquisition depended on the fact that they recurred. As Tomasello puts it:

> In general, if a child were born into a world in which the same event never recurred, the same object never appeared twice, and the adult never used the same language in the same context, it is difficult to see how that child – no matter her cognitive capabilities – could acquire a natural language. (p. 88)

TBLT in general, and studies of task repetition in particular, can reasonably make the same assumption. If they are important for the learners, the language learning tasks which help to constitute shared attentional frames in classrooms will also recur, and some of them will by their very nature – for instance, tasks such as regularly completing the weather chart, rehearsing different directions on familiar maps and plans (say of the school or the locality), re-organising the furniture in rooms or the layout of a shop or the placement of animals in a zoo, updating recommendations for equipment in the light of new products, and so on. What happens when tasks are repeated is therefore of interest.

Task repetition studies within TBLT research

Focussing on task repetition contrasts with most TBLT research in one key way. TBLT research has mainly concentrated on what happens when students engage with particular tasks or task types, or under particular task conditions. This has enabled exploration of a number of important issues, such as how particular task processes (such as negotiation for meaning sequences) might contribute to learning; whether tasks

can be designed to shape learners' language; whether the type of task (such as problem-solving tasks) or task conditions (such as presence or absence of planning time) influence the fluency, accuracy, or complexity of language, and by implication how these processes and types and conditions might shape acquisition. Research into these issues has tended to show that the use of tasks can shape learner language performance, and hence to some extent learning, and that different tasks and different task conditions tend to give rise to different degrees of fluency, accuracy, and complexity of language. However these studies have generally compared performances, either of different tasks, or else of the same task under contrasting conditions (for instance, as noted above, with or without pre-task planning). Less attention has been paid to two really central issues: firstly how far learners improve in handling particular tasks, and secondly how tasks might relate both to other classroom activities in the context of lessons and to out-of-school activities. These are crucial questions. Both are unavoidable for any approach which claims to make tasks central to teaching, and both lead to an interest in task repetition.

The first question, how far learners improve in handling particular tasks, bears on the central proposition of TBLT – that tasks are often selected for inclusion in programmes because learners need to be able to handle them. This kind of rationale makes tasks into what might be referred to as the 'target of learning', with adequate task performance becoming the criterion for successful learning. In other words, if one of the reasons for using tasks is because learners need to learn how to master them, then it is not enough for them to perform the tasks once: teachers and learners need to know whether as a result of working on a given task, they have improved their ability to handle them. This implies some form of repetition of the task, both in order to help learners master them, and also in order to enable us to assess whether they have improved. For instance, if a target task entails being able to participate as an applicant in job interviews, to explain safety procedures on an aircraft, to narrate historical events as a tour guide, or to engage successfully in seminars, then a one-off experience of the relevant task is not going to be adequate either for learning or for demonstrating learning achievements. The second question, how tasks might be used as central elements of classroom pedagogy, implies that they cannot be free-standing, and if this is the case they must play an important role in the coherence of lessons. This suggests that we need to know how their use can connect to other language learning activities, and more precisely, how learners and teachers themselves can make those connections.

It is these concerns which suggest the value of studying task repetition: we cease to think of tasks as being 'one-off' activities, but rather iterated as part of a sequence, something learners work on rather than do and discard. And if it is useful to see tasks within the context of a lesson, then maybe learners might sometimes have the opportunity of performing a given task or task type more than once – for instance, early

in a lesson to get a feel for the challenge, and then again later in the lesson after some background work, and later on, as homework at home (Van den Branden, personal communication). And if tasks are a 'target of learning', then it becomes possible to think in terms of learners working on tasks during a lesson or scheme of work in preparation for an exit task at the end of the learning sequence. The task then becomes the focus of teaching and learning, and in order to enable learners to develop their mastery, there might be several different types of task-based activities, leading to an assessment-type task.

This is the kind of thinking that has underpinned the interest in task repetition. It raises questions such as whether performance on a task at one time affects subsequent performance, and in particular which aspects of task performance might change on the second occasion, and to what extent. Other questions include how far collaborative and task-oriented strategies emerge or evolve as familiarity with the task increases; whether fluency, accuracy, or formal range change across iterations, and if so, whether this happens for everyone the same way; whether all aspects change, and if so whether this happens in parallel, or cyclically; how whole-class activities that take place between iterations might feed into subsequent task performance; and how far learners change their perceptions and understandings of tasks on subsequent encounters, and if so, in what ways.

Early research into task repetition

Early research into task repetition had more limited concerns than those outlined above, both in conception and in scope. Furthermore, findings were inconsistent. In one of the earliest published task repetition studies, Plough and Gass (1993) found that immediate repetition of a task increased the incidence of negotiation for meaning exchanges, suggesting a greater willingness among learners to focus on form to clarify meanings; however, they also reported that students were generally demotivated by having to repeat the task. Bygate (1996) reported a case study in which, by repeating a task after a three-day interval, a learner significantly improved their grammatical accuracy, their idiomaticity, and their lexis, and produced significantly more self-corrections. This suggested that on repeating a story from memory, learners were likely to shift their attention from concentrating on recovering the content of the story to considering aspects of formulation – in a sense consistent with the findings of Plough and Gass (1993). Similar effects were found by Lynch and Maclean (2000, 2001). In a series of six related case studies, using data recorded during what they called a 'poster carousel' activity, (an activity in which pairs of students prepared posters summarising different medical research studies and then talked about the poster to a sequence of different visitors), they found that through the iterations of the different visits,

irrespective of precise topic or level of proficiency, all six students improved the ways in which they formulated their talk – including pronunciation, grammatical accuracy, and lexical expressions.

However, studies have not all shown the same patterns of change. A larger scale study of a single repetition of a task after a 10-week interval (Bygate, 2001) showed that while fluency and complexity increased, accuracy was unaffected, suggesting that task repetition did not necessarily help direct attention to formulation but either helped in fluency of execution, or else led learners to complexify their talk. In addition, this study explored the possibility that as well as task-specific benefits, performing a task might lead to learners developing discourse strategies for managing the genre. With this in mind, students were also given five repeated encounters every 2 weeks with different exemplars of two distinct types of task (what has since come to be referred to as 'procedural task repetition' as in Kim et al., this volume) to see if this gave them an advantage when they carried out new tasks of the same types. The results however were null: task repetition seemed to have an effect when the same constellation of content material was involved, but if the content changed, learners seemed not to benefit from working with the task type. In a further related study Bygate and Samuda (2005) explored a subset of the 2001 data, focusing on the accuracy and complexity of four of the specific task repetitions that had taken place every two weeks. This inspection revealed clear increases in the complexity of students' language. Detailed analysis showed that this was due to increases in the amount of elaboration – especially the use of significantly more adverbials of place, time, manner, cause, and purpose. Whereas in the first iteration learners had simply given the bare bones of what had happened during the narrative, the second time around they filled out the circumstances of the story, along with more of the attitudes, motivations, and intentions of the different characters. This finding suggested that repetition could help learners by giving them the time and space to make the plot line more explicit, thus enabling them to improve the overall quality of their story. All the above studies involved adult learners. Pinter (2005, 2007), however, working with primary school learners of English, found increases in fluency, accuracy, and complexity of children's language when invited to re-do a task they had previously performed. In addition, when asked, children were generally very positive about having the opportunity to repeat the task, contrasting with the finding reported by Plough and Gass (1993).

Together these findings suggest three provisional conclusions. Firstly, all studies found significant differences in one or more aspects of performance between iterations. No studies reported null results when comparing two or more performances of the same task. This suggests that given particular conditions, on different iterations of a task, learners are likely to process the task differently, and that prior performance of a task is likely to affect how learners go about it the second time around. The likelihood that prior performance of a task changes how learners deal with a language task

is potentially very important, especially for teaching and for testing. A second conclusion from the early studies was that task repetition functioned similarly across different ages and indeed across distinct levels of proficiency: participants in Bygate's (1996 and 2001) and Bygate and Samuda's (2005) studies were randomly grouped to control for proficiency. Meanwhile Lynch and Maclean's (2000, 2001) reports presented case studies of speakers of very different levels of proficiency all nonetheless showing similar patterns of change. Thus there are grounds for tentatively concluding that changes across iterations in various aspects of language are likely to arise irrespective of proficiency level. A third provisional conclusion however is that although learners' language is likely to change across iterations, we cannot confidently anticipate whether this will occur predominately in terms of complexity, or accuracy, or fluency. These studies also oblige us to accept that unless learners see the challenge and possible payoff from repeating a task (which was the case with the children in Pinter's studies), they are likely to disengage from a task repetition cycle. And from a pedagogical perspective this raises important issues as to how task repetition sequences are framed by teachers in the classroom.

In brief, the picture emerging from the early studies on task repetition is that, provided they accepted the challenge, learners of all levels of proficiency are likely to benefit from task repetition by changing their focus and the resulting output. This hints at a significant learning dynamic that could underlie a sequence of task performances. However the picture also suggests the possibility that on any given repetition cycle, some learners may take the opportunity to complexify their language, some may hone their accuracy, and others may use their familiarity from the previous iteration to operate more fluently. That is, the focus of that underlying learning dynamic could be different for different learners at any one time. It is worth dwelling for a moment on the theoretical and practical significance of there being a possible learning dynamic along these lines.

Task repetition and learning in different fields

Much of the TBLT research to date has tended to concentrate on surface aspects of the learners' task performance, focusing on learners' language processing, generally in relation to some feature of the task design or the classroom conditions (such as presence or absence of planning). However the finding that there might be significant differences between performances of the same task by the same learner suggests that valuable learning could be taking place below the observable surface of the activity. The notion of the activity having 'an observable surface' points to the distinction between the observable qualities of the learners' discourse, such as its discourse structure, its pragmatic and cohesive features, its lexico-grammatical content, and its

complexity, accuracy, and fluency, on the one hand, and on the other, the underlying psycho-linguistic processes such as the strategic and purposeful conceptualisation, formulation, articulation, and monitoring of the intended communication (Kormos, 2006; Levelt, 1989). Models of language use, such as those of Levelt (1989) for speech and Bereiter and Scardamalia (1987) for writing, which bring together both underlying processes and surface qualities in this way, can be described as 'complex integrated models' (Deane et al., 2008). The term 'integrated' underscores the intention of the models to incorporate all relevant processes within the whole active skill.

Describing a project for developing diagnostic problem solving skills of medical professionals, Lajoie (2009) outlines the model relevant to their needs:

> There are different types of knowledge representations that are part of the overall problem-solving task. For instance, there is declarative knowledge (factual knowledge pertaining to a particular disease), conceptual knowledge (what one understands), and procedural knowledge (knowledge of what procedures or skills are necessary to test hypotheses and reach a final diagnosis). The overall cognitive model encompasses all of these competencies including contextual knowledge (knowing when to use and apply knowledge) and meta-knowledge (knowing what one knows or does not know). (p. 74–75)

As in many other domains, the area of interest involves the development of an integrated skill. The term 'integrated skill' is however intended to refer to the integral contribution of different levels of control, from the macro to the micro. Kahneman (2012) remarks:

> Learning high-level chess can be compared to learning to read. A first grader works hard at recognising individual letters and assembling them into syllables and words, but a good adult reader perceives entire clauses. An expert reader has also acquired the ability to assemble familiar elements in a new pattern and can quickly "recognize" and correctly pronounce a word that she has never seen before. (p. 238)

Learners need to develop the ability, firstly to handle the full hierarchy of skills, and then to improvise to handle new material both in comprehension and production. And for Lajoie (2009), tasks are relevant here both for the process of pedagogical design and for the teaching:

> In constructing a cognitive model of diagnostic reasoning, we start with selecting a task or case that represents the complexity of the domain. (p. 75)

If we accept that language use involves both surface features and underlying processes, then when learners repeatedly encounter the same or similar tasks, changes in performance may involve different aspects of the whole skill – including any of the various underlying processes, as well as any or all of the different aspects of surface production. Teachers and learners might then interpret the provision of repeated encounters

with a given task as providing a space for task learning, where learners can attend to different aspects of the language, its selection, and its production, in real time.

There seem to be parallels here with other types of learning. For instance, when learning a musical instrument, it is standard practice to work over and over again on specific pieces of music. Sometimes particular passages will be practised in isolation in order to get the fingering right, or else the tempo might be deliberately varied (Johnson, 1996); at other times the learner will be asked to play with slight variations of rhythm, and often the piece will be learnt by heart. Behind the repetitions, learning can be assumed to be taking place, in terms of improvements in the accuracy, the fluency, or simply the interpretation of the performances. The kinds of practice that sportsmen and women engage in is in some ways similar: they rehearse many times over a repertoire of moves they can expect to have to use during matches (Ericsson, 2009), and through the rehearsals different facets of learning are assumed to be taking place. Ericsson points out further that structured rehearsals in tutored contexts are likely to be far more effective than trying to learn simply from playing matches:

> The central assumption is that an individual's performance on a training task will vary as a function of focus of attention, type of strategy, and many other situational factors. Consequently, any type of deliberate practice is designed to maximise gains to performance by allowing the performer to be fully rested at the start of the activity. The performers should also be fully prepared for initiation of the task, be given immediate feedback from the outcome, and then be allowed to repeat the same or similar task with gradual modifications. Performing the task under these optimal conditions is more effective than performing the associated task only when it occurs within the natural context of performance. Hence, part of the practice is to gradually embed the trained task in its natural context with regular time constraints and less predictable occurrences. For example, imagine an amateur tennis player who misses a volley at the net. The play will go on until some time later a similar situation emerges unexpectedly with a similar problem for the player. Contrast this type of on-the-job learning with a session with a tennis coach. The tennis coach would set up situations where the player would stand at the end and be ready to execute the volley. With mastery of the easy volleys, the coach can increase the difficulty of the shots and eventually embed the volley shots into the rallies. It is easy to see that a few hours of this type of training would improve the player's volley more than 10s or 100s of hours of regular tennis play against other amateurs. (pp. 416–417)

The selection of – and then sustained practice on – a target task is intended to set up an opportunity to ensure the focused practice of contextually valid skills. The purpose is to structure the practice so that it is neither "mindless, routine performance" nor "playful engagement" (p. 417). He explains:

> The latter two types of activities would, if anything, merely strengthen the currently inadequate cognitive mediating mechanisms rather than modify them to allow increases in the level of performance. (p. 417)

The problem with both "mindless routine performance" and "playful engagement" is that they allow merely adequate performance to become automated at a point where the learner is satisfied with their level, making it difficult for them to progress any further. In contrast, "individuals aspiring to reach expert performance never allow their performance to be fully automated" (p. 417), a point echoed by Levelt (1978) when he remarked that the problem with audio-lingual approaches was that premature automation resulted in what he described as "collapsing the [skill] hierarchies" into frozen behavioural routines, blocking further development. Similarly Lajoie (2009) writing about an air force training project remarks:

> one dimension of expertise pertains to knowledge organization. As efficiency increases in a domain, elements of knowledge become more interconnected so that proficient individuals access coherent chunks of information versus fragmented knowledge. (p. 73)

In other words, across iterations of tasks, learners' underlying knowledge systems need to be gradually reorganised better to respond to the task demands, a point made many years earlier by Bruner (1960) in relation to school level mathematics teaching.

Research shows that the principle of avoiding premature automation and instead continuing to engage in deliberate focused practice applies to a very wide range of activities – tennis, piano playing, chess, athletics, ballet, typing, even darts. A further advantage of deliberate practice is that it avoids the risks that can ensue if practised in real life contexts: "when a surgeon performs surgery on a patient, this is not the time to try out an alternative method or focus on improving some particular weakness" (Ericsson 2009, p. 420).

A key component of successful learning in contexts using tasks has been found to be the presence of feedback associated with repeated practice (p. 423). Johnson and Jackson (2006) report a project in which they studied training programmes using tasks in which feedback was seen by the trainers as critical – pilot training, table tennis coaching, and coaching singers. Chatham (2009) describes the feedback facilities available to support military training exercises (made possible of course by what for applied linguists constitutes unusually generous levels of funding), in which the nature and sources of potentially critical errors in decision-making could be jointly identified, tracked down, and unpicked, openly by all ranks, for future reference. Lajoie (2009) reports on an Air Force training project the success of which she claimed was partly attributable to the fact that instruction and assessment were "interwoven", using a version of "dynamic assessment" whereby

> [w]hen students reach an impasse they get assistance that helps them move closer to their goals rather than being stopped in midstream while problem solving. Assistance is situated in specific contexts where learners can examine their strategies compared to experts. (p. 72)

The main implication here is that to enable learners to get the kinds of focused deliberate practice and ensuing feedback which Ericsson is urging, using tasks in and of themselves is not sufficient. For one thing, regularity is likely to be important if feedback is to be effective in building up expertise. Kahneman (2012) again notes that there are two basic conditions for acquiring a skill:

– an environment that is sufficiently regular to be predictable
– an opportunity to learn these regularities through prolonged practice (p. 240)

The potentially wide variety of language tasks available for use in classrooms brings with it the possibility that different tasks may create too wide a range of task demands, reducing the learning potential of the approach. Task repetition therefore offers a way of regularizing the task-based learning environment: first of all, reiterations of a task through a lesson or series of lessons reduces the risk of rapidly shifting from one task to another, instead helping to create moments in the programme where learners can concentrate extensively on a given task; and as a result, by working extensively on a given task, learners can be helped to develop a mental schema for its conceptual and pragmatic demands, become able to predict what the task involves, and thus better able to anticipate relevant language. The task environment becomes something manageable rather than a constant sequence of new challenges.

Additionally, task repetition may also be important in relation to the feedback that TBLT generates for learners. For one thing, the rich range of demands and opportunities for language use generated by tasks carries with it the risk that sequences of different tasks will offer limited opportunities for the re-use or application of feedback from previous tasks. In contrast repeated iteration of tasks could help to provide opportunities for learners to mobilise feedback from previous encounters in the context of upcoming iterations. Furthermore, often not all relevant feedback can be provided to a learner on a single iteration of a task. And even if it could be provided, all relevant feedback can usually not all be integrated into a learners' attention or memory on a single occasion. The complexity of tasks can mean too much new material in a given iteration, leading to learner overload (see also Hattie & Timperley, 2009). This clearly is a concern for pedagogy. Ensuring learners can encounter a given task on repeated occasions may create opportunities for progressive internalization of different aspects of the feedback, by enabling attention to be cyclically focused and re-focused. It might be useful at this point to consider the potential scope of task repetition for research and pedagogy.

Task repetition in language pedagogy

Since the notion of repetition has suffered a bad press for some time, being seen as boring, or associated with mechanistic reproduction of sentences, it is worth reiterating that the kind of repetition focused on in this volume is defined as the repeated performance of tasks that share some of the same pragmatic purpose or purposes and some of the same content. This definition is intended to include a number of slightly different ways in which meaningful task repetition can be used in the classroom. The various types all share an underlying theoretical and practical motivation for their use, centring on the ability of task repetition to promote the gradual integration of the various dimensions of relevant language knowledge into more fluent language use. Some of the different types of task repetition are summarised in Table 1.

Table 1. Some types of task repetition

Type 1	Repetition of task with exactly the same content in same arrangement (e.g., picture story, or instructing listener in arrangement of an array, completion or labelling of a map or diagram, provision of description or explanation) but with different interlocutor
Type 2a	Same array (list, diagram, room plan, model, map) but different arrangement
Type 2b	Same map or plan but different itinerary, or same picture paired with different contrasting picture
Type 2c	Same array (items for prioritising, recommending) with additions
Type 3	Chart or matrix with closed set of variations (e.g., weather chart)
Type 4	Information array with different interlocutors free to highlight different aspects (e.g., poster carousel)
Type 5	Schedule or matrix for series of interactions, description (e.g., interview schedule, survey), or extended work plan or project
Type 6	Transfer of material from oral to written or vice versa

Type 1 task repetition

The most basic type of repetition involves nothing more than the same content material repeated with the same overall purpose. This accounts for activities in which the precise same task (narrative, map task, comparison task, prioritising task, or other) is carried out a second time though possibly with a different interlocutor. In this type of repetition the speaker or writer covers the exact same overall content, although of course some of the material may be altered – omitted, rephrased, or supplemented by the learners – in relation to the first iteration. This type of repetition also accounts for the kinds of repetition built into Samuda's (2001) study in which the teacher asked

groups of students to report to the class the speculations that they had come up with so far about the likely owner of the 'things in pockets'. In this case repetition occurred in the interim reports provided to the teacher. After the interim reports, students went back to working in groups, where it could be expected that they would repeat once more the speculations that had been exchanged across the class.

Type 2 task repetition

A second type of task repetition involves tasks with the same overall purpose and the same array (such as the same map, same plan of a building, room, or of a location such as a park or zoo) but with a different itinerary, or a change in the arrangement of the array (the zoo animals are in different locations, or the furniture in the room has been moved around, or the plant display has been rearranged). Other minor changes could be made to an array while still maintaining the broad elements of task repetition. For instance, items of equipment to take on a hike could include an additional object or two, leading students to review the original prioritisation; or the original list of candidates for a position could be supplemented by two late applicants; a task to decide the best location of a factory on an area on a map could be followed by the request to locate a new building to house a builders' supplier's business (which would require the students to review the same geographical options for the new business in light of the previous siting). This type of repetition resembles Robinson's (2015) proposal for gradually complexifying a more simple entry task.

Type 3 task repetition

A third type of task repetition is represented by the daily teacher-class completion of the weather chart. Here a stable matrix is recycled from session to session, but since the weather varies from day to day, it is the generic structure rather than the precise information that is recycled daily. Yet the range of weather patterns is limited, and most of that will recur at some point or other during the annual cycle.

Type 4 task repetition

Task repetition can also occur where the content material is held constant but the interlocutors are free to focus on different aspects of it in light of their needs. This reflects the structure underpinning Lynch and Maclean's (2000, 2001) poster carousel activity. The poster creators stand by to explain aspects of the poster to other students who circulate around the room. Each 'visiting' student may well ask for clarification on different aspects of the poster: the overall content remains the same, but the focus of the interaction will vary from one visit to another.

Type 5 task repetition

'Partial' repetition of a task is a possibility that should also be taken into account: for instance, a job interview simulation task would typically involve recycling the same questions for each interviewee, but since the interviewees would respond differently to each question, the interviewers would have to adjust to the different replies, and offer different follow-up questions. Survey-type tasks are another example.

Type 6 task repetition

Finally, meaningful repetition can also occur when material that has been rehearsed orally or in writing is then reiterated or reviewed in another medium – the fruits of discussion work being written up, or written information being subsequently discussed orally. This type of repetition can also reasonably include transcription of student recordings, self-correction/self-editing, or summarising, and may well extend to include crosslinguistic reiterations, in which material handled in one language are then mediated through a second language (Norris, personal communication).

Although these various types of repetition differ, the repetitive dynamic can nonetheless be expected to have a similar role in helping the learner to integrate aspects of language into their communication. Elizabeth Gatbonton and Norman Segalowitz (for example 1988, 2005) have long been arguing for the value of repeating communicative activities in promoting fluency. However the argument can be taken further, since what is at issue through the various types of repetition enumerated above is the possibility of a unique contribution to learning which offers the possibility of reviewing, exploring, and elaborating language knowledge through iterations of activity on particular tasks. From this perspective the task can be seen as a stable 'workspace' which learners engage with, perhaps individually, in groups, and in whole class mode, and in which form and meaning can be extensively explored, and integrated into enactment of the task. This contrasts most strongly with the practice followed in some pedagogies of separating form-focused teaching from work on communication-oriented activities, an approach suggested explicitly for instance by Brumfit (1984). It also represents a distinct approach from that which would use tasks as a one-off self-contained activity at the end of a lesson. Further, the use of task repetition being suggested here offers a way of bringing a focus on form into TBLT which contrasts with the use of recasts or negotiation-for-meaning episodes. Whereas recasts typically occur on a one-off ad hoc basis in response to incidental communication problems arising during a task, the construct of task repetition sees a task as a rich site for exploring language possibilities, with learning occurring within a dynamic that develops across and between iterations. Thus it offers the possibility of each iteration enabling learners and/or teachers to review aspects of the task as well as potentially useful language

(whether previously familiar, or new), and in the process progressively integrate the language into the world of the task. In this way task repetition can set up the conditions to explore iteratively both the task and the language which might be relevant to it. I have suggested elsewhere (Bygate, 2006, p. 172) that one of the most important facets of task repetition may be the possibility of its enabling a 'constructive' or 'creative' engagement on the part of the learner with both the task and the language. Hence although the notion of task repetition starts from the idea of the exact same repetition of a task, the construct implies much more than this.

Pedagogical applications of task repetition

Exploration both of the task and of relevant language for the task implies working with tasks over time. This in turn means working with a given task both through and across lessons. If a task is worth mastering (which as noted above is often why it has been selected for inclusion in a programme) then it can naturally become the focal point of a lesson or scheme of work. This opens up the possibility of the task, or one version of it, being used in various phases of the lesson or scheme of work, for instance to introduce them; to raise challenges to the learners during the lesson; to create a focal point for exploration of relevant language; to set up writing phases; and of course as an exit point from the scheme of work. To this end, any of the different variations of task repetition identified above may be of value. Using tasks in this way also brings with it significant implications for task-based teaching and for research.

Perhaps one of the more far-reaching implications of the above is for tasks to be used to structure lessons, a type 5 variety of task repetition. This use of tasks would enable them to provide a sustained dual focus through the lesson – a focus both on the task and on relevant language – running in parallel, and interweaving in different ways at different times. This interweaving could be exemplified in the kind of lesson that can be observed in ESL contexts in UK secondary schools (see for instance Cameron, Moon, & Bygate, 1996). In one case, a series of woodwork lessons during which students were to make a wooden puzzle began with a verbal presentation of the woodwork task, accompanied by a diagram, and a list of steps that the students would be following – including tools to be used, procedures to be followed, and the safety precautions. The diagram and list of steps were accompanied by checks to ensure adequate understanding, with clarification where it turned out that the briefing involved new vocabulary. The students built their puzzles through a series of lessons, each lesson beginning with a reminder of the task briefing and the safety instructions, and a recapitulation of where the students had got to. During each lesson the teacher circulated, allowing students to ask questions and enabling him to comment and make suggestions. The task outcome was both a glass-topped puzzle (consisting of a silver

ball in a labyrinth) produced by each student, importantly for us, accompanied by a verbal account by the students of what they had produced and how they had done it. This scheme of work illustrates the iterative role of the task, as well as its dual focus. The task itself was worked on repeatedly over several lessons, and the dual focus involved engagement both with the practicalities of making the puzzle and with the relevant language needed to be able to mediate it and report on it. New language became relevant as new aspects of the task became clearer to the students. Interactions and specific questions, instructions, recommendations, explanations and descriptions were thus reiterated within and across lessons. Thus the task provided an overarching frame for the negotiation and renegotiation of the work that needed to be done, and to evaluate what had been done. As a whole this enabled firstly the interweaving of a focus on language and its use, and secondly the gradual exploration of the task and the language potentially relevant for its enactment.

There are of course numerous other ways of exploiting task repetition across lessons or schemes of work – the woodwork lesson is merely one illustration. However the point is that the use of task repetition can go well beyond the simple exact repetition of a task, to encompass numerous repetitions of all or part of the task, across one or more lessons. It also leads beyond the kind of approach (such as in the early proposals of Willis, 1996) which tended to limit the structure of lessons to a sequence made up of a pre-task phase, a task phase, and a post-task phase. Rather, the task can be allowed to take on a more central role.

This perspective broadens significantly the potential interest of task repetition to teachers as well to researchers. It raises questions such as what strategies teachers can use at different phases to negotiate the interweaving between task and language focus across whole or partial iterations of the task; how teachers can vary the focus between the task and its language potential; how the students' grasp of the task evolves through iterations; how their perceptions of the value of different elements of language change as they engage with the task; the range of different aspects of the language that can be illuminated through the different iterations, such as markers of modality, reference, time, conjunction, aspect, mood, and of course the deepening of lexical awareness; the extent to which the fluency, accuracy, and complexity of their language can develop through the various iterations of a given task; and the ways in which learners can collaborate in jointly construing the task and exploring the language for its mediation. These questions all bear centrally on the development of a rounded TBLT pedagogy, to which we hope this volume can contribute.

The scope of the volume

As can be inferred from the previous section, the pedagogy of task repetition is open to a wide range of research approaches. As the impossibility of mapping the real world in all its life-sized details makes clear, no single research methodology can fully render the human reality of an applied linguistic phenomenon – a principle from which task research is of course not immune. In any case it is probable that significant parts (if not all) of our research methodologies are complementary rather than mutually incompatible. With this in mind, the chapters in this volume bring together studies using methodologies informed by distinct theoretical frameworks. Two of the chapters in the volume are invited think-pieces (DeKeyser; Larsen-Freeman). Of the empirical chapters, the first six draw on cognitive theories (de Jong & Tillman; Kim, Crossley, Jung, Kyle, & Kang; Wang & Chen; Hu; Aubrey; Sheppard & Ellis), yet despite the fact that cognitive theories have had a relatively wide exposure in recent years, all six are somewhat unusual. The first three concentrate on relatively neglected aspects of language performance, lexical production (in the case of de Jong & Tillman, and Kim et al.), and discourse features (Wang & Chen), while the chapters by Hu, Aubrey, and Sheppard and Ellis are concerned with the impact on performance of implementation factors: the effect of inducing high or low criterion performances from groups (Hu); the impact of group, comparing NS-NNS versus NNS-NNS dyads (Aubrey); and the impact of introducing a post iteration stimulated recall procedure between two cycles of a task. The remaining five papers – exploring more explicitly the processes that can emerge from the task repetition dynamic – draw on socio-cultural frameworks (Lynch; Kobayashi & Kobayashi; Shintani) and on complex/dynamic systems theory (Nitta & Baba; Larsen-Freeman). Data for several of the studies come from lessons taught within an on-going language programme: ESP programmes in Lynch's chapter, adult EFL programmes in the chapters by Kobayashi and Kobayashi and by Nitta and Baba, and a beginners' programme for young children in Shintani's chapter. Thus, jointly the chapters that make up this volume also go some way towards reflecting the various ways in which task repetition can contribute to the development of TBLT.

In the opening chapter Robert DeKeyser reviews background research into task repetition within the skill acquisition literature. Echoing issues outlined in this introductory chapter, and our argument for a less tightly controlled approach to the concept of task repetition and to its use than is sometimes advocated, he reviews research on several themes relevant to task repetition, tending to lead towards a loosely controlled structuring of repetition. Themes include the distinction between procedural and declarative knowledge, and automatization. No less relevant to discussions of the use of task repetition across lessons or schemes of work, he notes the support in the literature for: (a) distributed practice, rather than repeated practice massed in a block

without intervening time intervals; (b) findings also in favour of the interleaving of different types of practice activity between instances of the target activity; and (c) for the repetitions themselves to be varied rather than being held constant. In a similar vein, he concludes by suggesting that even a relatively tightly defined construct such as automation may need to be seen as varying according to the domain of language, while the preferable degree of distribution of practice might be affected by whether the knowledge being practiced is procedural or declarative. The various trends that DeKeyser foregrounds are highly suggestive both for pedagogy and for research.

In their chapter, de Jong and Tillman investigate the extent to which speakers repeat language items in a repeated picture story, according to whether they are working under time pressure, and whether or not the repetition is immediate. They found that an increase in the incidence of repetitions of language items was associated with immediate repetition, and also with reduced fluency, but not with increased time pressure. Thus it was possible that language items were more likely to be repeated by less fluent speakers as a strategy to maintaining their language flow, and in repeated repetition contexts due to the relative recency of the previous iteration. In line with one of the themes of DeKeyser's opening chapter, these findings suggest how context can impact on patterns of processing.

Kim, Crossley, Jung, Kyle, and Kang use a collaborative task to explore the effects of task repetition and task complexity on learners' use of L2 lexis. The study adopts a 2×2×3 design to investigate the contrast in the language of simple versus complex tasks, in exact versus procedural repetition, across three repetitions – three exact repetitions of the same simple and complex tasks, and three procedural repetitions of similar simple and complex tasks. The authors report a positive relationship between amount of sophisticated lexis repetition of both exact and procedural repetitions of both the simple and complex tasks, with task complexity found to be a mediating factor.

While features such as fluency, accuracy, and complexity have been widely investigated, Wang and Chen's chapter investigates an aspect of task performance that has been the focus of much less study, namely the relative elaboration of learners' discourse. Using a simple repeated picture story design, the authors report more complex discourse on the second iteration of the story, in terms of amount of discourse produced, the amount of cohesion, and the quality of lexis.

The next three chapters explore ways in which the implementation of a task repetition sequence can impact on learners' language. In Scott Aubrey's study, the focus is on whether it makes a difference for learners to be paired with another non-native speaker (NNS) or with a native speaker (NS). Aubrey found that on the second iteration of the task, NNS-NS pairings yielded more language-related episodes (LREs), more LREs resolved with uptake, and a higher rate of complex or grammatical LREs. In contrast, the LREs of NNS pairings remained at the same level of frequency on the

second occasion, but with a drop in lexically oriented LREs. This study highlights the role that interpersonal factors can play in task-based communication.

In contrast to Aubrey's study, Hu investigates the impact of potential personal perspectives on task-based communication. Normally task-based communication is mediated by the extent to which the learners see the task as a relatively high or low stakes activity. Those with a high stakes orientation will engage a maximum in ensuring that they get their message across and that their interlocutor understands, while those with a lower stake orientation will be more ready to 'satisfice'. Thus two of Hu's experimental groups were given different follow-up tasks to the first task, the high criterion (HC) group a more demanding one than the low criterion (LC) group. The two groups then did a second 'repeat' task and the data analysed to see whether setting a high or low criterion had an impact on the performance of the second task. The HC group duly performed at higher levels of fluency and phrasal complexity on the second task, though lower levels of subordination complexity on the picture story task. Hu explains this in terms of the HC group's intention to ensure adequate comprehension, with the students shortening their utterances (reducing subordination) but increasing the precision of their narratives via phrasal complexity.

Sheppard and Ellis report a study that compares the impact of repetition on three repeated enactments of the same task and then on a new task, with one of the two groups in addition participating in a stimulated recall activity after the initial performance of the first task. They found that for both groups, structural complexity increased from the first performance and carried over to the new task. Accuracy stayed constant while fluency improved, but both fluency and accuracy declined on the new task, though were better than on the initial performance of the first task. The group that participated in the stimulated recall procedure showed greater fluency than the comparison group, though not on the new task. This study has at least two particular points of interest. The first is that it is one of the few studies that investigates the possibility of transfer of training from one task to a new one (sometimes referred to as 'procedural repetition' on the assumption that it is the procedures rather than the content of the task which might be internalised and transferred). Although some studies have found evidence of procedural transfer, support on the whole for this is limited, suggesting that learners internalise the content of the task more than its discourse procedures – which could be explained by the possibility that learners are often easily able to transfer their L1 discourse procedures, even if these are different in the L2. A second point of interest and possible discussion is the pedagogical value of using a stimulated recall procedure between repetitions of a task. It would appear from this study that some kind of pedagogical intervention between iterations can have an effect, a point that would be endorsed by others such as Lynch (this volume) or Samuda (2001). The question arises then whether a stimulated recall procedure is a definitive representative of this kind of intervention, or whether other teacher initiatives might

have different other kinds of impact. The negative impact on accuracy – a finding common to other (though not all) task repetition studies – is one for further exploration, since it seems unlikely that no learner would be able to take advantage of repetition to improve their accuracy (as indeed Bygate's, 1996 case study showed). It remains to be seen how, and under what conditions this might occur.

Several of the chapters show in different ways how task repetition can enable learners to explore more fully both the task and the language resources that can help to negotiate it, and indeed one of the chapters, by Tony Lynch, illustrates the same principle underlying the author's own pedagogic and research trajectory, which as his account of it shows, evolved dynamically through his career. In contrast to the other chapters in this volume, Lynch chooses to review a series of five of his own studies of classroom activities in order to draw general conclusions about the functioning of task repetition. Each of the studies was structured around a different classroom activity, each one designed to involve some form of what he terms 'enhanced repetition'. Four of the studies showed positive effects for the repetition cycle, while the fifth resulted in a null finding. Lynch takes the opportunity to reflect on why the fifth design was unsuccessful while the first four designs worked as expected. One key difference was that in the fifth study, the students re-interpreted the instructions, so that instead of responding to questions about their poster, they first gave an introductory speech, which they then followed up by reading the poster to the visitors. A second difference seemed to be a lack of appreciation of the value of the question-answer interaction which they had been asked to engage in. A third crucial difference seems to have been the lack of understanding of the value of the kind of repetition which the activity was intended to promote. Thus Lynch's chapter reminds us firstly of the importance of the internal design of tasks, but equally of the need for learners (and indeed teachers) to understand the rationale behind the tasks and their design, which to some extent echoes the argument underpinning Hu's study. The chapter also reminds us, however, of the fact that engaging in research is itself a dynamic learning activity.

Drawing on Lynch and Maclean's (2000) and (2001) research (see also Lynch, this volume), Kobayashi and Kobayashi contribute a chapter which investigates the classroom implementation of a poster carousel task. The authors use a socio-cultural framework to carry out a case study of a single group engaging in three consecutive question-answer sessions around the same poster. And like Shintani's chapter (see below), significantly the study reveals how the group adjusted their handling of the task through the iterations, drawing upon responses from their audience during earlier interactions to alter the way in which they made subsequent presentations and to adjust the ways in which they structured and managed the sessions. As noted above, the data analysis reveals how the repeated iterations enabled participants to explore both the task and the language, in this case enabling them to explore ways in which the task can be managed.

As in Lynch's chapter, Shintani explores a task as realised as part of an ongoing series of lessons. In this case the task is an input task jointly enacted by the teacher-researcher and the students, illustrating how a task can be realised in teacher-class mode, and not only in individual or group work formats. The task is a Type 2 version of task repetition (see Table 1 above) in which the same array is used repeatedly, but with different arrangements. Like Kobayashi and Kobayashi, Shintani draws on a socio-cultural framework to show how the task evolves as the learners gradually master the task and acquire the language. This seems a clear instantiation of the socio-cultural theoretic conception of tasks evolving as the participants themselves change, reminiscent of the phenomenon reported by Bruner in his (1983) study, in which the children bit by bit came to participate more in more in the formats, with the parent withdrawing to the point where the child managed the discourse event more or less entirely. Language acquisition and discourse participation go hand in hand.

Adopting a Complex/Dynamic Systems Theory (C/DST) approach, Nitta and Baba's chapter carries out a case study of the development of the L2 writing of two students through a series of tasks over 30 weeks of a taught programme. The two students showed contrasting patterns of development, one evidencing greater fluency, syntactic complexity and lexical sophistication than the other, which they relate to that student's more elaborated patterns of engagement and self-regulatory processes. The chapter provides a clear illustration of the ways in which a task repetition design can help to shed light on learners' learning processes, their engagement with the learning tasks, and on how these processes and overall engagement can relate to their language development.

The final chapter by Larsen-Freeman brings together some of the wider implications of using C/DST to explore the dynamics arising through repeated iterations of a task. Larsen-Freeman builds on the idea developed throughout the present chapter that learners constantly "construct their own learning paths" by adapting their language resources through different encounters with given tasks, which she prefers to refer to as 'iterations', making clear that through renewed encounters learners do not simply reproduce identical performances. Larsen-Freeman extends her argument to propose that C/DST enables us to reconceptualise the transfer of learning between contexts as *transformation*, arguing for an approach to pedagogy and research which recognises the agency and uniqueness of learners and of their personal construals that are intrinsic to their learning.

This is the first collection of papers published on the theme of task repetition, and if only from that perspective is almost certainly overdue. Issues arising for further discussion and exploration include the extent of the range of varieties of task repetition in classroom contexts; the degree to which fluency, accuracy, complexity and other aspects of performance might each develop on different iterations; the ways in which iterations reflect changes not only in the learners' language but also in their

understanding and mastery of the task; the ways in which teachers and learners can review and reflect on different iterations and as a result engage differently with further iterations of the task; and apart from exploring these various kinds of change, research might perhaps also enable us to better understand what it is that is being repeated. The variety of types of repetition, of contexts, of methods and theoretical orientations explored by the authors offers a basis – and suggests a potential – for substantial further work, and for the use and development of the construct within language programmes. I would like to repeat my appreciation to all the contributors for agreeing to participate in this collection, and for their enthusiasm and commitment during the preparation of the publication.

Envoi

Kahneman (2012) contrasts the differences between fast and slow thinking. Perhaps we should consider the possibility that the development of the ability to handle tasks, sometimes complex tasks, in a second language is a relatively slow, long-term process. Discussing a range of complex skills he notes:

> The acquisition of expertise in complex tasks such as high-level chess, professional basketball, or firefighting is intricate and slow because expertise in a domain is not a single skill but rather a large collection of mini skills. (p. 238)

It is possible that second language ability is indeed a collection of mini skills. It is also at least conceivable that the ability to handle tasks is also mediated through a collection of subskills. This is one perspective that could suggest the value of applying the concept of task repetition for the development of task-based language pedagogy. Along with the chapters contributed by the authors for this volume, this raises a rich range of empirical questions for future exploration in our various language classrooms.

References

Bereiter, C. & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Brumfit, C. J. (1984). *Communicative methodology in language teaching*. Cambridge: Cambridge University Press.

Bruner, J. S. (1960). *The process of education*. Cambridge, MA: Harvard University Press.

Bruner, J. S. (1983). *Child's talk*. Cambridge: Cambridge University Press.

Bygate, M. (1996). Effects of task repetition: Appraising the developing language of learners. In D. Willis & J. Willis (Eds.), *Challenge and change in language teaching* (pp. 36–46). London: Heinemann.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 23–48). Harlow: Pearson Education.

Bygate, M. (2006). Areas of research that influence L2 speaking instruction. In E. Uso-Juan, & A. Martinez-Flor (Eds.), *Current trends in the development and teaching of the four language skills* (pp. 159–186). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110197778.3.159

Bygate, M., & Samuda, V. (2005). Integrative planning and the use of task repetition. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 37–74). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.11.05byg

Cameron, L. J., Moon, J., & Bygate, M. (1996). Development of bilingual pupils in the mainstream. *Language and Education*, 10(4), 221–36. https://doi.org/10.1080/09500789608666710

Chatham, R. E. (2009). The 20th century revolution in military training. In K. A. Ericsson (Ed.), *Development of professional expertise* (pp. 27–60). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511609817.003

Deane, P., Odendahl, N., Quinlan, T., Fowles, M., Welsh, C. & Bivens-Tatum, J. (2008). *Cognitive models of writing: Writing proficiency as a complex integrated skill* (ETS Report RR-08–55). Princeton, NJ: ETS.

Ericsson, K. A. (2009). Enhancing the development of professional performance: Implications from the study of deliberate practice. In K. A. Ericsson (Ed.), *Development of professional expertise* (pp. 405–431). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511609817.022

Ericsson, K. A. (Ed.). (2009). *Development of professional expertise*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511609817

Ericsson, K. A., Charness, N., Feltovich, P. J., Hoffman, R. R. (Eds.). (2006). *The Cambridge handbook of expertise and expert performance*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511816796

Gatbonton, E., & Segalowitz, N. (1988). Creative automatization: Principles for promoting fluency within a communicative framework. *TESOL Quarterly*, 22(3), 473–492. https://doi.org/10.2307/3587290

Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *The Canadian Modern Language Review*, 61(3), 325–353. https://doi.org/10.3138/cmlr.61.3.325

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112. https://doi.org/10.3102/003465430298487

Johnson, K. (1996). *Language teaching and skill learning*. Oxford: Blackwell.

Johnson, K., & Jackson, S. (2006). Comparing language teaching and other-skill teaching: Has the language teacher anything to learn? *System*, 34, 532–546. https://doi.org/10.1016/j.system.2006.08.002

Kahneman, D. (2012). *Thinking, fast and slow*. Harmondsworth: Penguin.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, MJ: Lawrence Erlbaum Associates.

Lajoie, S. P. (2009). Developing professional expertise with a cognitive apprenticeship model: Examples from avionics and medicine. In K. A. Ericsson (Ed.), *Development of professional expertise* (pp. 61–83). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511609817.004

Levelt, W. J. M. (1978). Skill theory and language teaching. *Studies in Second Language Acquisition*, 1(1), 53–70. https://doi.org/10.1017/S0272263100000711

Levelt, W. J. M. (1989). *Speaking. From intention to articulation*. Cambridge, MA: The MIT Press.

Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. Special issue *Tasks in language pedagogy. Language Teaching Research*, 4(3), 221–250.

Lynch, T., & Maclean, J. (2001). A case of exercising: Effects of immediate task repetition on learners' performance. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 141–162). Harlow: Pearson Education.

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Chichester: Wiley.

Pinter, A. (2005). Task repetition with 10-year-old children. In C. Edwards, & J. R. Willis (Eds.), *Teachers exploring tasks in English language teaching* (pp. 113–126). Houndmills: Palgrave Macmillan.

Pinter, A. (2007). What children say: Benefits of task repetition. In K. Van den Branden, K. Van Gorp, & M. Verhelst (Eds.), *Tasks in action: Task-based language education from a classroom-based perspective* (pp. 131–158). Newcastle upon Tyne: Cambridge Scholars.

Plough, I., & Gass, S. (1993). Interlocutor and task familiarity: Effects on interactional structure. In G. Crookes & S. M. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 35–56). Clevedon: Multilingual Matters.

Robinson, P. (2015). The Cognition Hypothesis, second language task demands, and the SSARC model of pedagogic task sequencing. In M. Bygate (Ed.), *Domains and directions in the development of TBLT* (pp. 239–264). Amsterdam: John Benjamins. https://doi.org/10.1075/tblt.8.04rob

Tomasello, M. (2003). *Constructing a language. A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Willis, J. (1996). *A framework for task-based learning*. Harlow: Longman.

CHAPTER 1

Task repetition for language learning

A perspective from skill acquisition theory

Robert DeKeyser
University of Maryland

This chapter looks at task repetition in second language learning from the point of view of skill acquisition theory and related areas of psychology. It explores not only what concepts like procedural and declarative knowledge, automatization, and transfer have to offer the applied linguist, but also summarizes the literature on distribution of practice, interleaving versus blocking, and similarity in repeated practice, drawing on both the psychological and the fledgling second language acquisition (SLA) literature. In general, the literature shows support for distributed practice, for interleaving, and for variety in practice. For all these points there are exceptions, however, in particular in the SLA literature. This chapter discusses some of the possible explanations for these differences, and ends with some recommendations for where to focus future research: how automatization may vary with the domain of language, how distribution of practice may depend on whether the knowledge practiced is declarative or procedural in nature, and how research can help find the ideal context for bringing about transfer from one task to the other.

When applied linguists speak about repetition in the context of second language practice, this word almost inevitably conjures up images of drills, where repetitiveness is perhaps THE defining feature. On the other hand, when tasks are discussed in the context of SLA, and of task-based language teaching in particular, the image that comes to many people's mind most readily is one of authentic real-life tasks that are 100 percent communicative in nature and therefore cannot include much repetition. In that sense task repetition almost seems like a contradiction in terms. Neither 'repetition' nor 'task' have to be interpreted in such a radical manner, however. I will argue in this chapter that skill acquisition theory and other areas of cognitive psychology have much to contribute to a discussion of how tasks and repetition, both taken in a broader sense, cannot do without one another if second language skills are to be achieved.

Long (2015) states that "tasks are the real-world activities people think of when planning, conducting, or recalling their day" (p. 6). He contrasts this with "counterfeit

DOI 10.1075/tblt.11.01dek

tasks," which are "used to practice structures." Here I will use 'tasks' in a broader sense, including tasks that are designed to practice elements of language and that vary widely in how authentic they are. My use of the term, therefore, includes the way the word is used in discussions of task-supported teaching (Ellis, 2003; Samuda & Bygate, 2008). The latter define task as "a holistic activity, which engages language use in order to achieve some non-linguistic outcome while meeting a linguistic challenge, with the overall aim of promoting language learning, through process or product or both" (p. 69). This definition leaves room for activities that may appear much more didactic than the ones that fall under Long's definition; they may include tasks such as retelling a story a few times in an attempt to improve fluency.

Such activities and their repetition are far from limited to the classroom. They ARE representative of the real world, where we often make calls to several hotels to find the right room for the right night, relate a startling experience we had to several friends and relatives on the same day, approach several colleagues with the same problem we are trying to solve, and so on. This is an example of task-type repetition (Bygate, this volume); in other cases repetition can be internal to the task. This can occur naturally (as when the task is to moderate a debate and the moderator repeatedly asks people to voice their opinion or repeatedly summarizes what has been said so far). This distinction between task-type repetition and internal task repetition is fairly subtle, depending on how the task is delineated, of course. Small 'tasks' may be a repeated component in larger 'tasks' given to more advanced learners.

Less subtle perhaps is the distinction between task-as-workplan and task-as-process (Ellis, 2003, pp. 5–6; this volume): in any discussion about tasks we always have to keep in mind that the task as conceptualized and presented by the teacher can take different shapes as it unfolds over time through the interaction of the participants.

Our task as applied linguists, meanwhile, is multi-faceted. Long (2015) distinguishes five areas in need of research about tasks. The two most relevant for my purposes here are "classifying and sequencing target and pedagogic tasks" and "transferability of task-based abilities" (p. 373). In the field of second language acquisition (SLA), progress in the first area has been hampered by a lack of common definitions of dependent and independent variables, and in the second one we have seen very little research so far; most outcome measures are taken from a task almost identical to the one practiced (and often no delayed testing on the same task is even carried out). Researchers in cognitive and educational psychology, on the other hand, have amassed a substantial body of research on distribution of practice, transfer, transfer-appropriate processing, and the nature of practiced knowledge, but their research is seldom about tasks as complex as engaging in second language (L2) discourse, and certainly not as complex as the non-'counterfeit' L2 tasks.

Contributions from skill acquisition theory

As the purpose of this chapter is to analyze what skill acquisition theory and related areas of cognitive psychology have to contribute to the issue of task repetition in SLA, a brief presentation of the core concepts in that area is in order, and the first thing that comes to mind when talking about skill acquisition theory is the distinction between declarative and procedural knowledge. It is often said that declarative knowledge is "knowing that" and procedural knowledge is "knowing how." A slightly less concise but perhaps clearer definition is that declarative representations (whether episodic or semantic) are objects of thought, and procedural representations provide the (cognitive) actions to work upon these objects (e.g., Gade, Druey, Souza, & Oberauer, 2014); these cognitive actions are often called (if-then) production rules. In English as a second language, an example would be knowing what the word 'run' means or knowing verbs like 'run' take an –s in the third person singular of the present tense (declarative knowledge) versus being able to use that word and that ending in the production and comprehension of English (procedural knowledge). It should be pointed out, however, that the opposite of declarative knowledge is simply non-declarative, and that procedural is one subcategory of that (along with habituation/sensitization, simple classical conditioning, and capacity for priming).

An overlapping distinction is made between rapid and slow learning mechanisms (Henke, 2010). Even just following a single train of stimulation, long-term potentiation can be induced in the hippocampus (the main area responsible for declarative, in particular episodic, learning), while the basal ganglia (largely responsible for procedural learning) require stimulation that is repeated over many days in order to reach potentiation. Equally important is that the elements of declarative knowledge are not blended into an inseparable or unitized representation, but are accessible individually or in relation to each other, while procedural memories can consist of multiple and related elements, but these elements are usually blended into an inseparable or unitized representation and cannot be reactivated individually. As a result, the elements of declarative memory can be reactivated through many routes, and this flexibility permits the inferential use of memories in new retrieval situations; procedural memory is less flexible. This has important implications for the task specificity and very limited transferability of procedural knowledge, which we will return to later.

A central idea of skill acquisition theory is that declarative knowledge can lead to procedural knowledge and eventually to highly automatized procedural knowledge. Several misconceptions have clouded understanding of this issue in applied linguistics circles. First of all, as I have pointed out before (DeKeyser 2009, 2015, 2017), it can be misleading to state that declarative knowledge 'turns into,' 'changes into,' or 'gets converted into' procedural knowledge, because that terminology seems to imply for many that an existing chunk of knowledge gradually changes its nature (which

would also mean 'moving' from one area of the brain to another), which is not the case. Instead, declarative knowledge plays a causal role in the acquisition of procedural knowledge: declarative knowledge helps the learner to engage *(repeatedly)* in the mental behaviors that allow the building of procedural knowledge (after a number of repetitions of those behaviors). Secondly, declarative knowledge, besides words, does not only consist of rules; both rules and examples (and prototypes, collocations…) are part of declarative knowledge and can be drawn on in the proceduralization process (e.g., Anderson, Fincham, & Douglass, 1997; Gatbonton & Segalowitz, 1988). Finally, and most importantly perhaps from an applied point of view, repetition does not mean 'drill and kill.' Any kind of problem-solving behavior is subject to proceduralization, including solving the problem of how to put a specific type of sentence together, and this proceduralization requires repeated engagement in the task. For reasons of experimental control, of course, empirical research on practice or automatization in language learning tends to be very narrowly focused on a few rules that are applied many times (e.g., Bird, 2010; de Jong, 2005; DeKeyser, 1997; Rodgers, 2011; Suzuki & DeKeyser, 2017), but this does not imply that the proceduralization and automatization of rules cannot take place in a more varied and communicative context. There is nothing incompatible then between skill acquisition theory and tasks, even in the sense of task-based language learning.

Most of the literature on skill acquisition has been behavioral in nature, and differences in the nature of knowledge as a result of practice have been inferred from error rates, reaction times, and interference with/from competing tasks (e.g., Ericsson, 2006; Ericsson, Krampe, & Tesch-Romer, 1993; Logan, 2002; Newell & Rosenbloom, 1981; Rosenbloom & Newell, 1987). These studies and reviews have documented how performance, as measured in error rate or reaction time, improves as a power function of the amount of practice. In the last couple of decades, however, a number of studies and meta-analyses have accumulated that provide evidence from neuroimaging. They have demonstrated (mostly with functional magnetic resonance imaging, fMRI, but also with positron emission topography, PET) how the areas that show increased activation during a specific task change as a function of practice. Interestingly, these areas appear to be very much the same for both verbal and non-verbal cognitive tasks and for visuomotor skills (for reviews, see Chein and Schneider, 2005; Henke, 2010; Hill and Schneider, 2006; Squire & Wixted, 2011).

Contributions from other areas of cognitive psychology

Two other topics that have received quite a bit of attention in cognitive psychology are essential to our purpose here: the substantial literature on distribution of repeated

practice (the spacing issue and the interleaving issue) and studies about kinds of repetition (degrees of similarity).

The issue of distribution of spacing has a long research tradition. Cepeda, Pashler, Vul, Wixted, and Rohrer (2006) conducted a large meta-analysis of this literature and concluded that, generally speaking, distributed practice is better than massed practice, but the ideal distribution depends on the delay between practice and testing. More specifically, they posit an ideal ratio of inter-session interval and retention interval ranging from about 10 to 30% (e.g., 1 to 3 days between sessions if the posttest is going to come 10 days later). A more recent meta-analysis (Rohrer, 2015) used much stricter criteria for the selection of studies and concluded that distributed practice is better, provided that the test delay is at least a month. At least three different explanations have been proposed, but the most likely seems to be the study-phase-retrieval account (Toppino & Gerbier, 2014), which argues that when there is more spacing between practice sessions, then retrieval is more difficult and overlapping elements are strengthened, which leads to greater abstraction and better retrieval. Of course, if the interval is too wide, then too much forgetting will occur between sessions for the practice to be effective.

A related but different issue is that of interleaving versus blocking different elements in practice. While the spacing effect issue pertains to list learning, and other forms of deductive learning, the interleaving issue applies to inductive learning. In category learning, in particular, there has been a debate about whether it is better to show blocks of examples of the same category rather than to interleave examples from the different categories to be learned (for instance, if the goal is to learn to distinguish pictures by Picasso from pictures by Matisse, is it better to present them in large blocks or to alternate them?). Kornell and Bjork (2008) found that interleaving worked better in their experiments, which they explained with their discriminative-contrast account: interleaving helps learners discriminate the critical features of the items, because it forces them to focus on differences, while blocked presentation only draws attention to commonalities. The advantage of interleaving has been replicated in a number of studies since (e.g., Birnbaum, Kornell, Bjork, & Bjork, 2013; Kang & Pashler, 2012; Wahlheim, Dunlosky, & Jacoby, 2011). More recently, however, Zulkiply and Burt (2013) showed that this advantage of interleaving only extends to low-discriminability categories and that high-discriminability categories benefit more from blocked exposure.

Finally, there is the question of similarity in repeated practice tasks. Research in cognitive psychology, not surprisingly, has shown that variability in practice is good for generalizability. More surprising is the finding that generalizing to condition X after practice under various conditions that did not include X may lead to better performance on task X than practice with task X itself. Kerr and Booth (1978), for instance, conducted a study where one group of participants practiced tossing beanbags at a

goal that was exactly 3 feet away, while the other group practiced tossing the bags at goals located at 2 and 4 feet away. The latter group did better on the posttests, even for the 3-feet distance, which they had never practiced.

As we will see in the next section, for the time being SLA research has yielded more ambiguous findings on the issues reviewed here.

Contributions from SLA research

In SLA research, the issue of distribution of practice has recently drawn attention too, along with work on the similarity of repetition. While research in cognitive and educational psychology has been largely in favor of distributed practice, as shown above, in SLA research the findings have been less clear. Bird (2010) found distributed practice to be best for the learning of past tense use in ESL, and Nakata (2015) showed an advantage for spacing in ESL vocabulary learning. Serrano (2011; Serrano & Muñoz, 2007), however, found that for an English course as a whole, more concentrated teaching (25 hours of instruction per week during five weeks) was more effective than the same number of hours distributed over three, four, or seven months (the dependent variables covered a wide range, from listening and reading performance to measures of accuracy, fluency, and complexity in oral production). Suzuki and DeKeyser (2016, 2017), in a study on the learning of Japanese morphology, found no difference for different amounts of spacing between practice sessions, but did instead detect an aptitude-treatment interaction in the sense that the massed treatment drew more on participants' memory capacity and the distributed treatment more on analytic ability. Given the different focus and the different time scales of these studies, it is hard to pinpoint the reasons for the differences in their findings and to come to any generalizations at this point. In Serrano's study there was no delayed testing, which could certainly explain the advantage of massed practice in their study. It could also be that more complex content (as in that study and the one by Suzuki and DeKeyser) benefits more from massed practice than the paired-associate learning or the dichotomous form choice found in most other studies.

When it comes to similarity of repeated practice tasks, available SLA research is even less in agreement with the psychological literature. Osthus (2014) studied the learning of three rules of Macedonian L2 grammar under blocked and interleaved conditions, and he did not replicate the findings reported above that interleaving becomes more important as discriminability decreases. In this study the blocked and interleaved conditions always yielded the same results.

Even less expected findings were obtained in a series of experiments by Carpenter and Mueller (2013) on orthographic-to-phonological mapping in French L2 (eight rules had to be induced from exposure to 64 new words). Whether blocking was

manipulated within or between participants, whether they learned 4 words per rule or 15, and whether the test format was recall or multiple-choice, the blocked condition always did better.

At this point it is unclear why neither of these studies found an advantage for interleaving. It may be that the rules and patterns to be learned should be considered to fall into the high-discriminability category, but that seems unlikely, given that the morphosyntactic rules in Osthus (2014) were of three different levels of complexity. It seems more likely that the short duration of these studies (without delayed testing) favored the blocked condition (analogously to what we have learned about massed practice), or that the advantage of interleaving is less likely to be obtained for the aural stimuli used in these experiments than for the visual stimuli used in all the other experiments on interleaving.

The question of similarity in repeated practice tasks has also yielded rather complex findings. On the one hand, de Jong and Perfetti (2011) found similarity to be beneficial. They had participants tell a story three times, at increasing speed, following the 4/3/2 minute model (Arevart & Nation, 1991), and repeated telling of the same story benefited fluency more than telling different stories, even if the posttest was yet a different story. Thai and Boers (2016), however, found that this advantage obtained for fluency only, and that repetition without speeding up, while having (smaller) advantages for fluency, managed to avoid the negative impact on accuracy that they observed for the speed-up condition. De Jong and Tillman (this volume) acknowledge this point, but the fact remains that variability was disadvantageous for learning when the outcome variable was fluency, a result diametrically opposed to what we saw in the Kerr and Booth (1978) bean tossing study cited in the previous section. The most likely explanation for the discrepancy, of course, is the nature of the task, but what exactly is it about the task that made the most consistent, narrow practice best for transfer to a new task? It may have to do with the nature of what needs to be transferred and how that is best learned. De Jong and Tillman show convincingly that when the task remained the same, participants repeated many elements, not only lexical ones, but more importantly strings with the same part-of-speech structure. These repetitions correlated with fluency gains. Repeatedly using these abstract structures, and hence their proceduralization (forming production rules through practice that draws on the corresponding declarative knowledge) and automatization (fine-tuning and strengthening of these production rules), happens more easily when the task remains the same than when the story to be told varies each time, yet the abstract structures being practiced this way are eminently transferable to other stories.

This leads us to the all-important issue of transfer. Researchers working in the area of skill acquisition theory have stressed the specificity of practice, and hence the very limited direct transfer of procedural knowledge from one task to another. When transfer seems to take place across more widely differing tasks, it is really transfer via

declarative knowledge, or, more precisely, what happens is that procedural knowledge is used to strengthen/expand declarative knowledge, which is more amenable to new tasks, and can then lead to new forms of procedural knowledge for those tasks (see esp. Singley & Anderson, 1989). In other words, for procedural knowledge to be useful, it has to be transferable, either through direct formation of new production rules (near transfer, i.e., to very similar tasks in very similar contexts, e.g., using the same politeness formulas to order from different menus) or via declarative knowledge (far transfer, i.e., to tasks that are similar, but in a more abstract way, e.g., using similar syntactic constructions in different business meetings). The research in De Jong and Perfetti (2011) and De Jong and Tillman (this volume) suggests that the procedural knowledge that can lead to near transfer is at a certain level of abstraction (strings of n words that have the same part-of-speech structure), that is, clearly more abstract than specific words or word strings, but perhaps not as abstract as grammar or discourse rules that could be used in any task. The findings in Fukuta (2016) also point in that direction: participants who repeated the exact same narration showed more attention (as assessed through stimulated recall) to syntax and less to content in the second than in the first performance; they also did better in terms of accuracy and lexical variety. Participants who narrated a different, but similar story did not show these effects (neither group was told to focus on form, but both the previous experience of the learners, university students in Japan who had learned English for many years, and the picture description task probably contributed to focus on form in both groups). Transfer to more widely different tasks may require that the execution of a task lead to an awareness of structure at a more abstract and hence more generalizable level.

That being said, the question still remains what the ideal nature of the procedural knowledge generated by the repetition of a specific task is. Research in educational psychology has emphasized the principle of transfer-appropriate processing (TAP) (Bransford, Brown, & Cocking, 1999; Martin-Chang & Levy, 2005, 2006; Morris, Bransford, & Franks, 1977; Roediger, Gallo, & Geraci, 2002), and SLA theory has drawn on it to some extent in the last decade or so (e.g., De Ridder, Vangehuchten, & Seseña Gómez, 2007; Lightbown, 2007; Spada & Lightbown, 2008), but very little empirical research has made it central to its design.

The basic idea of TAP is very simple: what learning condition will lead to the best posttest performance depends on the congruence between what is stressed during learning and what is in focus during testing. In Morris et al.'s (1997) study, for instance, both the learning and the testing stages had variants that stressed the semantics of words and others that stressed the rhyming properties. Martin-Chang and Levy (2006) had participants learn new words under two conditions: in a story during context training and on a computer screen during isolated word training. They found that words trained in isolation are remembered longer and read faster when presented in isolation at testing time compared to words trained in context.

What this means for the learning of procedural knowledge rather than for the declarative word learning in their study is harder to see and depends on what degree of transfer is required. For near transfer, what matters is that the procedural knowledge learned be very similar to that needed for the new task. The implication is that task sequences should recycle the relevant production rules wile varying the task, in particular by increasing its complexity. Often, however, the transfer has to stretch further, and then it is important that the practice tasks make the learners aware of the elements of declarative knowledge (e.g., the grammar rules or pragmatic routines involved) that will be required for new tasks (and that may subsequently be proceduralized in slightly different ways through practice with the new tasks). That may be the reason why Sheppard (2006, quoted in R. Ellis, 2016, p. 420) found that providing learners with input and feedback between performances led to improvements in complexity, accuracy, and fluency in the repeat performance. Such interventions could help learners form more stable declarative representations of what they practiced narrowly in the preceding performance.

In SLA, as in psychology, empirical research drawing on the TAP concept has been primarily on vocabulary. Trofimovich (2005), for instance, showed that L2 learners over-rely on context-specific information available from spoken input (e.g., the speaker's voice) as they take advantage of auditory word priming: such overly specific memories are obviously not transfer-appropriate. Stengers et al. (2016), in a study on idiom learning, once more show how narrow transfer can be. Based on the TAP concept, they hypothesized that stressing the semantics of idioms (by visualizing their literal meaning) would be better for recall of meaning and that stressing form by having students copy the idioms would be better for recalling exact form. This hypothesis was not borne out however, probably, as the authors themselves speculate, because the practice format still was not close enough to the format of the posttest (fill-in-the-gaps).

One SLA study that did address TAP in grammar learning is Spada, Jessop, Tomita, Suzuki, and Valeo (2014). They compared isolated form-focused instruction followed by practice with integrated form-focused instruction. They found little difference between the two on oral and written posttests, but a tendency for the isolated group to do better on the written task and for the integrated group to do better on the oral task. They see this as possible evidence for the importance of TAP, but acknowledge that their design (pre-test differences and lack of a control group) makes the data hard to interpret. Similar findings were reported by De Ridder, Vangenuchten, and Seseña Gómez (2007), who used a very wide variety of outcome measures in all linguistic domains, but which are hard to interpret in their case because of the lack of random assignment and the possibility of differences in motivation between their two treatment groups, one using a task-based approach and the other a traditional communicative curriculum.

At any rate, if a sequence of tasks incorporates much repetition from one task to the next, the exact nature of that repetition is crucial, both for transfer from one task to the next and for eventual transfer to the real world. The crux consists in determining how exactly the task similarity has to be understood. Skill acquisition theory says it is all about the similarity of production rules from one task to the next and about having a declarative equivalent to bridge the bigger transfer gaps, but other factors may play a role too: "future research will have to determine which dimensions of psychological similarity (e.g., whether the learners' intentions, feelings, etc., are important, or whether only linguistic contexts are important) are relevant to the establishment of automaticity that is transferable to new situations" (Segalowitz, 2003, p. 402).

As we saw above, the task-as-process as engaged in by the learner can be different from the task-as-workplan as conceived by the teacher, and it is precisely these various psychological factors that determine what exactly the learner pays attention to and learns as a result. Does the learner see the task as something to get over with or as something to learn from? If it is something to learn from, then for what: an exam, a job, travel plans? Attention will be focused differently depending on these goals, and also on the affective factors associated with them, ranging from anxiety to exuberance. When a task is repeated, then, the impact of these affective and motivational factors will be felt even more: the learner can focus on what can be learned from previous performances of a (similar) task that could bring him/her closer to these goals and the associated feelings, and going through this learning experience repeatedly with that mindset can make a big difference in the end.

In this sense, task-based language teaching has a big motivational advantage: if tasks are selected as a function of specific learners' needs, whether personal or professional, and sequenced as a function of difficulty, it is more likely that the learner will distill from the individual performances the procedural and declarative knowledge that seems most likely to be useful in future tasks, whether it be in class or in the real world.

Meanwhile, however, SLA research on the effects of task repetition has been somewhat limited so far, and has yielded rather divergent findings about what effects can be expected in terms of accuracy, fluency, and complexity, as several contributions to this volume make clear. It has also been somewhat insulated from relevant research in various areas of psychology. In order for this research to become more useful to practitioners, whether they adopt a task-based or a task-supported approach, it needs to incorporate the variables mentioned here and find out how they interact with task repetition, itself a variable concept.

What should be next

Starting with aspects of repetition itself, it is clear that the SLA literature has painted a somewhat different picture from that found in psychology. As shown above, the psychological literature has shown a clear benefit of distributed practice, while the results in SLA so far are more ambiguous, even with delayed testing. One explanation for this may be the more procedural, even automatized, nature of the knowledge tested in those SLA studies that did not find an advantage for distributed practice. It is important then, that we make this distinction, and more in particular that we have more studies about distribution of practice for proceduralization and automatization of grammar and pronunciation, not just vocabulary learning. Also, form-meaning connections and form-form connections may require different amounts/kinds of practice. Few questions could be of such obvious practical importance, while also holding considerable theoretical interest, as the psychological research shows. The practical question can be asked at several levels, from distribution of number of hours of L2 per week, to distribution of specific tasks and distribution of specific structures. All of this may miss the mark, however, if what matters is repetition of the procedural knowledge to be practiced and if the practice/tasks to be distributed are not defined in such terms. The same task-as-workplan may lead to different tasks-as-process, depending on the level of proficiency or how far the learner has evolved in the acquisition of a specific structure; this implies the use of different forms of knowledge (declarative, procedural, automatized) and hence potentially a difference in the ideal distribution of practice. Therefore, we need research on distribution of practice from this perspective.

Moreover, as Suzuki and DeKeyser (2016) show, distribution of practice may interact with individual differences in working memory and language-analytic ability. Research on individual differences has finally begun to investigate interactions with treatments head-on in the last decade or so (for overviews see DeKeyser, 2012; Vatz, Tare, Jackson, & Doughty, 2013), but interaction with distribution of practice has not been on the agenda so far and seems worth investigating.

Secondly, as Norris (2009) already pointed out, individual and contextual variables will also have an impact on how tasks are received, and this will certainly be the case for task repetition. Different cultures have different expectations about repeating and recycling, and individuals certainly have widely differing needs for and tolerance of (specific kinds of) repetition. We should not lose these variables out of sight while we focus on the skill acquisition perspective.

The most central mandate for future research on task repetition, however, is to think about what procedural knowledge the task sequence is meant to develop. What structures should be taken to what stage of skill development? How will one task pick up where a similar one left off in that respect? How will we make sure that enough

generalizable declarative knowledge is developed for transfer to tasks that are a bit more different? From my point of view, this perspective has been underrepresented in thinking about the repetition and sequencing of tasks and similar activities.

References

Anderson, J. R., Fincham, J. M., & Douglass, S. (1997). The role of examples and rules in the acquisition of a cognitive skill. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 23(4), 932–945.

Arevart, S., & Nation, P. (1991). Fluency improvement in a second language. *RELC Journal*, 22, 84–94. https://doi.org/10.1177/003368829102200106

Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics*, 31(4), 635–650. https://doi.org/10.1017/S0142716410000172

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory and Cognition*, 41(3), 393–402. https://doi.org/10.3758/s13421-012-0272-7

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn. Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Bygate, M., & Samuda, V. (2005). Integrative planning through the use of task-repetition. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 37–74). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.11.05byg

Bygate, M. (this volume). Introduction. In M. Bygate (Ed.), *Learning language through task repetition* (pp.1–25). Amsterdam: John Benjamins.

Carpenter, S. K., & Mueller, F. E. (2013). The effects of interleaving versus blocking on foreign language pronunciation learning. *Memory and Cognition*, 41, 1–12. https://doi.org/10.3758/s13421-012-0291-4

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

Chein, J., & Schneider, W. (2005). Neuroimaging studies of practice-related change: fMRI and meta-analytic evidence of a domain-general control network for learning. *Cognitive Brain Research*, 25(3), 607–623. https://doi.org/10.1016/j.cogbrainres.2005.08.013

de Jong, N. (2005). Can second language grammar be learned through listening? An experimental study. *Studies in Second Language Acquisition*, 27(2), 205–234. https://doi.org/10.1017/S0272263105050114

de Jong, N., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, 62(2), 533–568. https://doi.org/10.1111/j.1467-9922.2010.00620.x

de Jong, N. & Tillman, P. (this volume). Grammatical structures and oral fluency in immediate task repetition. In M. Bygate (Ed.), *Learning language through task repetition* (pp.43–73). Amsterdam: John Benjamins.

DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition*, 19(2), 195–221. https://doi.org/10.1017/S0272263197002040

DeKeyser, R. M. (2009). Cognitive-psychological processes in second language learning. In M. Long & C. Doughty (Eds.), *Handbook of second language teaching* (pp. 119–138). Chichester: Wiley-Blackwell. https://doi.org/10.1002/9781444315783.ch8

DeKeyser, R. M. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning*, 62(Suppl. 2), 189–200. https://doi.org/10.1111/j.1467-9922.2012.00712.x

DeKeyser, R. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition. An introduction* (pp. 94–112). London: Routledge.

DeKeyser, R. (2017). Knowledge and skill in SLA. In S. Loewen & M. Sato (Eds.), *Handbook of instructed second language acquisition* (pp. 15–32). London: Routledge.

De Ridder, I., Vangehuchten, L., & Seseña Gómez, M. (2007). Enhancing automaticity through task-based language learning. *Applied Linguistics*, 28(2), 309–315. https://doi.org/10.1093/applin/amlo57

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Ellis, R. (2016). Focus on Form: A critical review. *Language Teaching Research*, 20(3), 405–428.

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 683–703). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511816796.038

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363–406. https://doi.org/10.1037/0033-295X.100.3.363

Fukuta, J. (2016). Effects of task repetition on learners' attention orientation in L2 oral production. *Language Teaching Research*, 20(3), 321–340.

Gade, M., Druey, M. D., Souza, A. S., & Oberauer, K. (2014). Interference within and between declarative and procedural representations in working memory. *Journal of Memory and Language*, 76, 174–194. https://doi.org/10.1016/j.jml.2014.07.002

Gatbonton, E., & Segalowitz, N. (1988). Creative automatization: Principles for promoting fluency within a communicative framework. *TESOL Quarterly*, 22(3), 473–492. https://doi.org/10.2307/3587290

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews Neuroscience*, 11, 523–532. https://doi.org/10.1038/nrn2850

Hill, N. M., & Schneider, W. (2006). Brain changes in the development of expertise: Neuroanatomical and neurophysiological evidence about skill-based adaptations. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 653–682). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511816796.037

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, 26(1), 97–103. https://doi.org/10.1002/acp.1801

Kerr, R., & Booth, B. (1978). Specific and varied practice of motor skill. *Perceptual and Motor Skills*, 46, 395–401.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science*, 19(6), 584–592. https://doi.org/10.1111/j.1467-9280.2008.02127.x

Lightbown, P. (2007). Transfer appropriate processing as a model for classroom second language acquisition. In Z. Han (Ed.), *Understanding second language process* (pp. 27–44). Clevedon: Multilingual Matters.

Logan, G. (2002). An instance theory of attention and memory. *Psychological Review*, 109(2), 376–400. https://doi.org/10.1037/0033-295X.109.2.376

Long, M. (2015). *Task-based language learning*. Oxford: Wiley-Blackwell.

Martin-Chang, S. L., & Levy, B. A. (2005). Fluency transfer: Differential gains in reading speed and accuracy following isolated word and context training. *Reading and Writing*, 18, 343–376. https://doi.org/10.1007/s11145-005-0668-x

Martin-Chang, S. L., & Levy, B. A. (2006). Word reading fluency: A transfer appropriate processing account of fluency transfer. *Reading and Writing*, 19, 517–542. https://doi.org/10.1007/s11145-006-9007-0

Morris, D. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 16, 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning: Does gradually increasing spacing increase vocabulary learning? *Studies in Second Language Acquisition*, 37(4), 677–711. https://doi.org/10.1017/S0272263114000825

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Lawrence Erlbaum Associates.

Norris, J. (2009). Task-based teaching and testing. In M. H. Long & C. J. Doughty (Eds.), *Handbook of language teaching* (pp. 578–594). Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781444315783.ch30

Osthus, P. (2014). *Is interleaved practice better than blocked practice for aurally presented grammar rules? A comparison of online- and laboratory-based methodologies* (Unpublished qualifying paper). University of Maryland.

Rodgers, D. M. (2011). The automatization of verbal morphology in instructed second language acquisition. *IRAL*, 49, 295–319. https://doi.org/10.1515/iral.2011.016

Roediger, H. L. I., Gallo, D. A., & Geraci, L. (2002). Processing approaches to cognition: The impetus from the levels-of-processing framework. *Memory*, 10(5/6), 319–332.

Rohrer, D. (2015). Student instruction should be distributed over long time periods. *Educational Psychology Review*, 27, 635–643. https://doi.org/10.1007/s10648-015-9332-4

Rosenbloom, P., & Newell, A. (1987). Learning by chunking: A production system model of practice. In D. Klahr, P. Langley, & R. Neches (Eds.), *Production system models* (pp. 221–286). Cambridge, MA: The MIT Press.

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. Houndmills: Palgrave MacMillan.

Segalowitz, N. (2003). Automaticity and second languages. In C. Doughty & M. Long (Eds.), *Handbook of second language acquisition* (pp. 382–408). Oxford: Blackwell. https://doi.org/10.1002/9780470756492.ch13

Serrano, R. (2011). The time factor in EFL classroom practice. *Language Learning*, 61(1), 117–145. https://doi.org/10.1111/j.1467-9922.2010.00591.x

Serrano, R., & Muñoz, C. (2007). Same hours, different time distribution: Any difference in EFL? *System*, 35, 305–321. https://doi.org/10.1016/j.system.2007.02.001

Sheppard, C. (2006). *The effects of instruction directed at the gaps second language learners noticed in their oral production.* (Unpublished Ph.D. dissertation), University of Auckland, New Zealand.

Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard University Press.

Spada, N., Jessop, L., Tomita, Y., Suzuki, W., & Valeo, A. (2014). Isolated and integrated form-focused instruction: Effects on different types of L2 knowledge. *Language Teaching Research*, 18(4), 453–473. https://doi.org/10.1177/1362168813519883

Spada, N., & Lightbown, P. M. (2008). Form-focused instruction: isolated or integrated? *TESOL Quarterly*, 42(2), 181–207. https://doi.org/10.1002/j.1545-7249.2008.tb00115.x

Squire, L. R., & Wixted, J. T. (2011). The cognitive neuroscience of human memory since H.M. *Annual Review of Neuroscience*, 34, 259–288. https://doi.org/10.1146/annurev-neuro-061010-113720

Stengers, H., Deconinck, J., Boers, F., & Eyckmans, J. (2016). Does copying idioms promote their recall? *Computer Assisted Language Learning*, 29(2), 289–301. https://doi.org/10.1080/09588221.2014.937723.

Suzuki, Y., & DeKeyser, R. (2016). Exploratory research on L2 distributed practice: An aptitude-by-treatment interaction. *Applied Psycholinguistics*, 38(1), 27–56.

Suzuki, Y., & DeKeyser, R. (2017). Effects of distributed practice on the automatization of L2 morphosyntax. *Language Teaching Research*, 21(2), 166–188.

Thai, C., & Boers, F. (2016). Repeating a monologue under increasing time pressure: Effects on fluency, complexity, and accuracy. *TESOL Quarterly*, 50(2), 369–393. https://doi.org/10.1002/tesq.232

Toppino, T. C., & Gerbier, E. (2014). About practice: Repetition, spacing, and abstraction. *Psychology of Learning and Motivation. Advances in Research and Theory*, 60, 113–189. https://doi.org/10.1016/B978-0-12-800090-8.00004-4

Trofimovich, P. (2005). Spoken-word processing in native and second languages: An investigation of auditory word priming. *Applied Psycholinguistics*, 26(4), 479–504. https://doi.org/10.1017/S0142716405050265

Vatz, K., Tare, M., Jackson, S. R., & Doughty, C. J. (2013). Aptitude-treatment interaction studies in second language acquisition. In G. Granena & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 273–292). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.35.11vat

Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: an investigation of mechanisms, metacognition, and aging. *Memory and Cognition*, 39(4), 750–763. https://doi.org/10.3758/s13421-010-0063-y

Zulkiply, N., & Burt, J. S. (2013). The examplar interleaving effect in inductive learning: moderation by the difficulty of category discriminations. *Memory and Cognition*, 41(1), 16–27. https://doi.org/10.3758/s13421-012-0238-9

CHAPTER 2

Grammatical structures and oral fluency in immediate task repetition

Trigrams across repeated performances

Nel de Jong and Philip Tillman
Vrije Universiteit Amsterdam / Independent researcher

In this study we examine to what extent words and grammatical structures are re-used when a speaking task is repeated with the same content (i.e., specific task repetition). We examine this re-use, which has been argued to support proceduralization and fluency development (N. de Jong & Perfetti, 2011), under both constant and increasing time pressure, and we investigate the correlation between re-use and fluency. The analyses are performed not only on individual words but also on trigrams, which are sequences of three words (e.g., *the red car*; here: lexical trigrams) or three parts of speech (e.g., DET ADJ NOUN: POS trigrams), to capture grammatical structure. Thirty-nine adult ESL speakers completed repeated retellings of one to three picture stories. One group followed the 4/3/2 procedure (Nation, 1989), which involves three iterations with gradually increasing time pressure; for the other group the available time was constant. The extent of re-use of words and grammatical structures across task iterations was calculated using cosine similarity with tf-idf weighting (Manning, Raghavan, & Schütze, 2008), which adjusts for the frequency of words or trigrams, both within an iteration and across iterations and speakers. It was found that immediate task repetition had a strong effect on re-use at the level of individual words and trigrams, but increasing time pressure did not. The relationship between re-use and fluency was variable, but showed higher re-use for speakers struggling with fluency. We conclude that, if fluency development is to be stimulated by re-use of words and grammatical structures, it can be done with specific task repetition, whether under increasing time pressure or not.

Introduction

Repeating speaking tasks can help second language (L2) learners improve their task performance and ultimately it may support their language development. Task repetition has been shown to support oral fluency, accuracy, and structural complexity in

DOI 10.1075/tblt.11.02jon

the repeat performance (e.g. Ahmadian & Tavakoli, 2011; Bygate, 2001; Nation, 1989; Sample & Michel, 2014). Following Bygate (2001), this is often attributed to speakers shifting their attention from message content to formulating the message. Task repetition can also lead to language development because it allows learners to re-use particular words and grammatical structures or use more advanced language items, and thus make their future retrieval faster and less effortful (N. de Jong & Perfetti, 2011; Kim & Tracy-Ventura, 2013), supporting the development of fluency and perhaps also complexity.

Despite the growing body of research showing the relationship between task repetition and task performance, the exact nature and extent of the re-use of words and grammatical structures remains unclear. One reason is that most studies only report measures of structural and lexical complexity of each performance, such as the number of clauses per AS unit or a measure of lexical variety. But these measures do not show, for instance, whether a similar value for a structural complexity measure was based on the same or different types of clauses, or whether a similar lexical variety score was based on the same or different words. A second reason is that it is commonly assumed that when a task assignment is repeated, task execution is similar as well. But this is not necessarily the case: speakers may choose to add, leave out, or change information, although different tasks allow this to different extents (personal narratives, for instance, are more open to change than picture story retellings). Establishing the extent of similarity of task execution is therefore important for interpreting task performance: if there is indeed a strong similarity between iterations, this similarity is a likely explanation for any increases in fluency, accuracy, or complexity; but if there is less similarity, a different explanation is necessary. Without an analysis of the similarity of task execution, explanations of task repetition effects remain speculative. The first aim of this study is therefore to examine the extent to which immediate task repetition leads to re-use of words and grammatical structures.

Task repetition in this study is examined in the context of the 4/3/2 procedure (Boers, 2014; N. de Jong & Perfetti, 2011; Nation, 1989), an activity that has long been used in language courses. In this activity, students speak for four minutes about a given topic, change partners to repeat their monologue in three minutes, and repeat it again to a different partner in two minutes. In a study by de Jong and Perfetti (2011), focusing on oral fluency development, one group of L2 learners performed this activity on three occasions and a comparison group did the same activities but with a different topic each time (i.e., repeating the task type but not the task content). Results showed that oral fluency increased across iterations for the content-repetition group but that it was variable for the non-content-repetition group. In addition, the content-repetition group maintained their increased fluency on a posttest with new content. An analysis of the lexical overlap between retellings showed that the content-repetition group repeated more words than the comparison group, even though most

of these words were not clearly related to the specific content of the task. De Jong and Perfetti attributed the development of oral fluency in the content-repetition group to proceduralization of linguistic knowledge (explained in more detail below), and not simply to faster lexical retrieval. They therefore called for a further analysis of the speakers' use of grammatical structures in relation to fluency, which is the second aim of this study.

To systematically examine the effects of both immediate task repetition and increasing time pressure (as in the 4/3/2 procedure) on the re-use of words and grammatical structures, in this study we analyze trigrams. Trigrams are sequences of three words (referred to here as lexical trigrams, e.g., *the blue car*) or three parts of speech (POS trigrams; e.g., 'determiner adjective noun'). Trigrams (as well as bigrams, 4-grams, etc.) have been used in second language research to analyze lexical bundles and formulaic sequences in the speech of native speakers (e.g. Biber, Conrad, & Cortes, 2004; Martinez & Schmitt, 2012; Simpson-Vlach & Ellis, 2010) as well as second language learners (e.g. Chen & Baker, 2016; De Cock, 2004; Lauttamus, Nerbonne, & Wiersma, 2007; McCarthy et al., 2009; Paquot & Granger, 2012; Thai & Boers, 2016). Although most trigrams do not represent complete structural units, they have been shown to have important grammatical correlates (Biber & Conrad, 1999; Biber et al., 2004). Analysis of the re-use of trigrams across task performances can provide insight into how much is repeated beyond the level of individual words. Not only does such an analysis provide a better understanding of the effects of task conditions on the fluency of task performance, it also shows whether task repetition is likely to support the development of oral fluency and transfer to new task types and new content with respect to the use of grammatical structures.

Immediate task repetition

To date, studies of task repetition have found mixed results as to its effect on task performance. De Jong and Perfetti (2011), for instance, showed that fluency increases across iterations in the 4/3/2 procedure, while de Jong (2012) found no changes in syntactic complexity. Bygate (2001), on the other hand, found that syntactic complexity, but not accuracy, was higher when a task type (interview or personal narration) was repeated with the same content than with new content, and fluency was lower when content was repeated in an interview. Yet another pattern of results was presented by Ahmadian and Tavakoli (2011): on a repeated film retelling, they found increases in syntactic complexity and fluency but not accuracy. Finally, Sample and Michel (2014), in a study with young learners performing a spot-the-difference task, found an increase in fluency but no significant differences in syntactic complexity, lexical complexity or accuracy. In sum, these studies have mostly shown effects on

fluency and syntactic complexity but not accuracy, but there is considerable variability, even within studies.

This variability may have to do with a range of factors, including the type of task repeated and the time interval between iterations. First, as for the type of task, it is important to distinguish between, on the one hand, the materials and instructions given to the speakers, which Ellis (2003, p. 5), following Breen (1989), calls the 'task-as-workplan', and on the other hand the speakers' interpretation and execution of the task-as-workplan, which is referred to as the 'task-as-process'. When speakers are asked to repeat a task, they are asked to repeat the task-as-workplan. But in fact, the question that is of interest to teachers and researchers is to what extent and in which respects repetition of the task-as-workplan, or elements of it, results in repetition of the task-as-process. When speakers repeat an interview or personal narrative (e.g. Bygate, 2001; N. de Jong & Perfetti, 2011), the task content is relatively unconstrained, and speakers are able to change the content of their task performance, for example, speaking about dogs the first time and cats the second time. This variability is likely to reduce the extent of repetition of the task-as-process. Other studies have used more constrained tasks, such as film retellings (Ahmadian & Tavakoli, 2011) or picture story retellings (N. de Jong, 2012), which limit the content that can and needs to be included. In addition, it is important to distinguish between repetition of the type of task and repetition of the content of a task. The current study examines task repetition using picture story retellings, which are relatively constrained in terms of content, and compares task-type repetition (retellings of different picture stories) to specific task repetition (retellings of the same picture story).

A second explanation for the variability in findings of task repetition studies is that they have differed with respect to the time interval between iterations. Bygate's (2001) study of task repetition involved intervals between the original and repeated performance of 10 weeks, while Ahmadian and Tavakoli (2011) chose a one-week interval and Gass, Mackey, Alvarez-Torres, and Fernández-García (1999) chose two to three days. In the 4/3/2 procedure, however, the interval is much shorter, up to a few minutes at most. We refer to this interval here as immediate repetition. The choice of interval most likely affects how and to what extent performance is affected, but to examine this assumption further, it is necessary to consider first what it is exactly that speakers do when they perform and repeat a task.

What speakers do when they repeat a task

To fully understand the effects of task repetition manipulations, it is essential to have a framework for describing what speakers do during their performance of a task. From a cognitive perspective, oral production can be broken down into stages: message

generation (conceptualization), lexical and (morpho)syntactic formulation, and articulation (Levelt, 1989, 1999). At different points before and after articulation, speakers monitor their speech. For smooth performance, processes in different stages need to operate concurrently, which requires a certain level of automaticity. Although conceptualization is not automatic in most respects, most processes in the formulation and articulation stages are automatic in native speakers. Crucially, they are not necessarily automatic in L2 learners. In fact, L2 knowledge often starts out as non-automatic, declarative knowledge and later may become proceduralized and automatized with repeated practice (see also Anderson, 1993; DeKeyser, 2009; Lyster & Sato, 2013).

Based on Levelt's model, Skehan (2009) argues that increased demands on message generation (conceptualization), including quantity of information to be expressed and certain types of planning, may lead to increased lexical and structural complexity. But since non-automatic processes require attention, and assuming that attentional capacity is limited, Skehan also argues that pressure on formulation, for example, time pressure or the need for infrequent lexis, may negatively affect formulation, lowering accuracy and fluency. In contrast, pressure on formulation can be eased by structure and planning, including task repetition. In short, task performance is affected by task demands and task conditions, but also by the level of proceduralization and automatization of a speaker's L2 knowledge.

The 4/3/2 procedure, which is investigated in this study, combines two different task conditions: immediate task repetition (of type and content) and time pressure. Task repetition is expected to affect task performance at the levels of conceptualization, formulation, and articulation. First, demands on conceptualization are highest in the first iteration, when content is generated, while demands are lower on the second iteration, when most or all content is already available. A second effect of task repetition is priming (see McDonough & Trofimovich, 2009). Priming occurs when recent activation of lexical and grammatical items makes them more readily available. Thus, priming decreases demands on formulation, which in turn supports performance. Since most priming effects are considered to be relatively short-lived, however, they will mostly be involved in immediate repetition, that is, repetition with intervals of no more than a few minutes. Finally, although repeated articulation has not shown priming effects in native speakers, for whom articulation appears to be overlearned (Francis, Corral, Jones, & Sáenz, 2008), it may still affect language learners, particularly at less than advanced levels.

The attentional resources that become available as a result of task repetition could be used to increase complexity, accuracy, and/or fluency. But how exactly these resources are used likely depends to a large extent on the individual speaker's choices: while one speaker may focus on accuracy, another may value complexity, and yet another may want to work on both. Furthermore, as Palotti (2009, p. 596) rightly argues, "linguistic complexity grows when this is specifically required by the task and

its goals, and not for the sake of it" (p. 596); the same could be said about accuracy. This suggests that without a clear need for complexity and accuracy, and without a speaker's specific focus on these aspects of performance, the benefit of repetition may be limited to fluency. This is consistent with Bygate's (2001) finding that in his control group, which did not receive any training between the original and repeat interview, the effect of task repetition after 10 weeks was limited to fluency. Also, Ahmadian and Tavakoli (2011) found that both fluency and complexity, but not accuracy, benefited from repetition of a film retelling within a week. Because the benefits of immediate task repetition have been shown fairly consistently to support fluency, fluency will be the primary focus of the present study.

The other task condition that is, by definition, characteristic of the 4/3/2 procedure is time pressure. When time pressure is imposed on a repeat performance, it appears that speakers have a choice between two strategies. One strategy is to simply leave out information, which is, in fact, what Nation (1989) found to occur most in a 4/3/2 procedure. Another strategy is to rely on words and syntactic structures that are highly available due to their high frequency, low complexity, or recent use (i.e., priming). Oral fluency is not likely to be affected by the former strategy but could be expected to increase as a result of the latter strategy, because words and structures that have become more readily available can be re-used. The two strategies affect similarity between iterations differentially: if speakers leave out information, similarity is suppressed, but if they rely on re-using words and syntactic structures from the original performance, similarity is high. Which option is chosen, if not both, is likely to depend on any specific instructions pertaining to these strategies as well as on individual preferences.

The present study examines the combination of immediate task repetition and increasing time pressure. The effects of repetition and time pressure may conflict or support each other, depending on the strategy chosen to deal with time pressure. Tentative support for this assumption is presented by Ahmadian and Tavakoli (2011). In their study, task repetition raised fluency while the absence of time pressure decreased it, and when both were present, they cancelled out each other's effect. They argued that when repeating a task, whether under time pressure or not, their participants chose to re-use language items, which had become more readily available. They did not, however, provide any data to show this was what their participants did.

Like the Ahmadian and Tavakoli (2011) study, most studies on task repetition to date have compared iterations of a task in terms of complexity, accuracy, and fluency, but few have examined to what extent iterations were similar to each other. There are a few notable exceptions. First, Bygate and Samuda (2005), in a reanalysis of Bygate's (2001) data, found that the lexico-grammatical changes between two film retellings on average amounted to only 7% of the total number of words in the corpus. Nevertheless, there were substantial increases in the amount of information content,

evaluations, or summaries. Second, de Jong and Perfetti (2011), looking for an explanation for an increase in fluency after three 4/3/2 training sessions, found that under the content repetition condition the number of words used in all three iterations was higher than under the non-repetition condition. Although this finding in itself is not unexpected, it was striking that this pattern was not limited to content-specific words: general words such as *good*, *really*, *friend*, and *remember* were also repeated more. With respect to task performance, de Jong and Perfetti found that the number of reused words correlated with fluency gains across iterations within the 4/3/2 procedure. This suggests that there was a lexical effect on fluency, whether due to lexical retrieval (formulation) or articulation. This lexical effect was not found for fluency development from pretest to posttest, however, as very few words used in the posttests had also been used in the training sessions. Therefore the authors argue that pretest-to-posttest fluency gains could not be attributed to improvements in lexical retrieval but rather to improvements at other levels, most likely morphological and syntactic levels.

Two studies examining the relationship between immediate task repetition, increasing time pressure, and task performance were reported by Boers (2014) and Thai and Boers (2016). In Thai and Boers's study, 10 adolescent EFL learners in Vietnam performed a version of the 4/3/2 procedure, talking about their favorite movie. They did so under two time conditions: decreasing time (three, two, and one minute, respectively) and constant time (two minutes each time). Results showed that fluency improved most in the decreasing-time condition, and structural complexity and accuracy improved only in the constant-time condition. Neither group showed any change in lexical sophistication. Crucially, verbatim repetition, counted as the percentage of words occurring in repeated word strings, was 70% in the second iteration and 85% in the third iteration in the constant-time condition, and as high as 80% on the second iteration and 91% on the third in the decreasing-time condition. These findings are supported by Boers (2014), who did a similar study but with ten adult ESL learners in New Zealand. Both of these studies show, in line with Skehan's (2009) expectations and Ahmadian and Tavakoli's (2011) findings, that time pressure limits the benefits of task repetition for accuracy and complexity.

The two studies by Boers and Thai are an important first step in examining the similarity between repeated performances. There are three important limitations to their methodology, however. First, both studies counted n-grams of variable length (i.e., the longest stretches of repeated words that were used), some of which were up to 30 words and longer. Although this certainly gives an indication of the amount of re-use, it does not show whether parts of those repeated stretches of speech also occur elsewhere in the same iteration. For example, if the 7-gram *fly to the waterfall in his house* (Thai & Boers, 2016, pp. 391–392) is used in both iterations, but *in his house* also appears elsewhere in the iteration, this would not be counted as a repeated stretch (or it would need to be counted twice). A second limitation is that the authors

calculated the number of all words in repeated n-grams over the total number of words used in both iterations. In other words, all words contributed equally to the calculation of similarity. It may be desirable, however, to take into account the frequency of the repeated words: repetition of a highly frequent function word such as *the* may be considered less informative than repetition of a less frequent content word such as *waterfall*. A more systematic analysis to compute similarity between texts can be found in the field of Information Retrieval (discussed below), by using a fixed n-gram length and by giving lower weights to more frequent words and trigrams. A final limitation in the analyses of Boers and Thai is that they only analyzed n-grams of exact word forms (e.g., *goes*, *met*, and *couldn't*). This conflates the re-use of words, inflections, and grammatical structures. Analyses of both single words (unigrams) and trigrams, as well as both exact words and syntactic categories, will be reported in the current study to provide a broader picture of re-use across iterations.

Measuring similarity between iterations

The extent of the similarity between task iterations can be represented using different methods, depending on what is counted and how similarity is computed.

What to count: Similarity at the level of the word and above

Re-use between retellings was calculated by de Jong and Perfetti (2011) for single words (lemmas), while Thai and Boers (2016) used uninterrupted word strings (n-grams). These n-grams provide information above the word level; they can be strings of two (bigrams), three (trigrams), four or more words (for examples, see Table 1). n-grams are often used in corpus linguistics to examine multiword expressions, as in the body of research by Biber and Hyland and their colleagues on the use of lexical bundles in academic writing (e.g. Biber et al., 2004; Hyland, 2008). In Natural Language Processing (NLP), n-grams are used for a range of applications, such as communication systems, author identification, and sentiment extraction (Jurafsky & Martin, 2008). What this research shows is that certain words often co-occur with certain other words, a phenomenon which varies, for example, by modality and genre (lexical bundles) and by language user (author identification).

As Biber, Conrad, and Cortes (2004) argue, n-grams can capture some grammatical structure, including verb phrase fragments (e.g., distinguishing active and passive, and yes/no and wh-questions), dependent clause fragments (e.g., *wh*-clauses, *if*-clauses, and *to*-clauses) and noun phrase or prepositional phrase fragments (e.g., noun phrases with a postmodifier, or prepositional phrases). While bigrams provide some information on word combinations, trigrams and 4-grams capture somewhat

more of the grammatical structure. 5-grams can capture structure over longer distances, but they become quite specific to a particular text and are thus less likely to be re-used. Since trigrams can capture grammatical structure and because some of the utterances in our data are quite short, in the present study we chose to use trigrams for our analyses.

Table 1. n-grams for the sentence "The red car is driving too fast"

n-gram length	n-gram instances	Number of n-grams
unigrams	the, red, car, is, driving, too, fast	7
bigrams	the red, red car, car is, is driving, driving too, too fast	6
trigrams	the red car, red car is, car is driving, is driving too, driving too fast	5
4-grams	the red car is, red car is driving, car is driving too, is driving too fast	4
5-grams	the red car is driving, red car is driving too, car is driving too fast	3

Trigrams of exact word forms, as used by Thai and Boers (2016), capture information about their morphology (in the same way as the n-grams in Table 1), but to explore the re-use of grammatical structure further, it is possible to use trigrams at the more abstract levels of lemmas (base words and all of their inflections) or syntactic categories (e.g., noun, verb, preposition; following NLP conventions, we refer to these as part-of-speech trigrams, or POS trigrams). While the lemma trigrams for the sentence "The red car is driving too fast" are *the red car, red car be, car be drive, be drive too, drive too fast*, the POS trigrams for the same sentence are DET ADJ N, ADJ N AUX, N AUX PART, AUX PART ADV:INT [intensifier], PART ADV:INT adv.[1] Since POS trigrams represent words as their syntactic categories, they reflect the grammatical structure of the utterances more directly than lemma trigrams and lexical trigrams. In this study, we chose to analyze trigrams at the two extreme levels of abstraction: trigrams of exact words and POS trigrams.

How to compute similarity: Using frequency to estimate informativeness

Unigrams and trigrams, both lexical and POS, provide an inventory of words and structures used in a task performance. Similarity between iterations can then be calculated in different ways, with methods ranging from more straightforward and intuitive to more sophisticated. The most intuitive calculation is the Jaccard Index (see Manning et al., 2008). This is the number of words or n-grams that two iterations (referred to as 'documents' in NLP) have in common (the set intersection) divided by

1. ADJ = adjective; ADV = adverb; ADV:INT = intensifying adverb; AUX = auxiliary; DET = determiner; N = noun, PART = participle

the number of words or n-grams that occur in one or both of the iterations (the set union); see Table 2 for an example. There are two extreme cases: if two documents have no words in common the index is zero; if they consist of exactly the same words the index is 1. Thus the Jaccard index is a value ranging from 0 to 1, with zero being not similar at all and 1 being identical (although the words may occur in a different order and with different frequencies).

Table 2. Calculation of the Jaccard Index of two utterances (here considered documents) based on unigrams

Utterance	Unigrams	Set union (all unigrams)	Set intersection (common unigrams)	Jaccard index
The red car is driving too fast	the, red, car, is, driving, too, fast	the, red, car, is, driving, too, fast, blue, slowly [9]	the, car, is, driving [4]	4 / 9 = .44
The blue car is driving slowly	the, blue, car, is, driving, slowly			

The Jaccard Index, while being conceptually simple and reasonably accurate in many cases, has an important limitation, in that it does not take into account the frequency of words. Whether a word occurs once or many times in a particular retelling, it is counted only once. This means that a retelling about cars, which contains the word *car* 100 times but uses *bike* only once can, in principle, have a high Jaccard similarity with a retelling about a bike, which uses *bike* 100 times and *car* only once. This is not desirable. On the other hand, a retelling in which *car* is used 200 times might still be very similar to one in which it is used 100 times. Therefore a calculation that takes into account the relatively frequency of each word would be desirable.

Another frequency-related issue in an intuitive calculation such as the Jaccard index is that words that are used by many speakers, including most function words, are treated in the same way as words used by few speakers. If most speakers use *car*, this word does not distinguish the speaker's retelling from other speakers' retellings. But if only one speaker uses *red*, this word is very informative: this particular speaker's retelling is different from other speaker's retellings. To distinguish between speakers it is the infrequent words that are the most informative.

In summary, it is informative to use the frequency with which words are used within and across retellings (referred to in NLP as term frequency and document frequency, respectively). To do so, each word can be assigned a weight representing its informativeness. This is most easily implemented not with the Jaccard index, but with a different measure of similarity: cosine similarity (Manning et al., 2008).

How to compute similarity: Cosine similarity with tf-idf weighting

Cosine similarity first represents each iteration as a list of words and their frequency, called the term frequency list (see Table 3). Next, the iteration (or 'document') is represented by a vector, which consists of the frequencies for each word. Cosine similarity then is the cosine of the angle between the two vectors (see Figure 1 for an example). To illustrate this, Table 4 shows five very short documents, the vectors of which are presented in Figure 1. The cosine of the angle between document 1 and document 2 can be computed as follows:

$$\cos\theta = \frac{v_1 \cdot v_2}{||v_1||\,||v_2||} \quad , \; v_1 \cdot v_2 = v_{1,1}v_{2,1} + \ldots + v_{1,n}v_{2,n} \quad , \; ||v_1|| = v_1 \cdot v_2$$

where

$$v_1 = (v_{1,1}, v_{1,2}, \ldots, v_{1,n})$$
$$v_2 = (v_{2,1}, v_{2,2}, \ldots, v_{2,n})$$

where $v_{1,n}$ and $v_{2,n}$ represent the nth component of vectors 1 and 2, respectively.

Table 3. Sample 'documents' and term frequency lists and vectors

Document	Term frequency list	Vector
The red car is driving too fast	{(the, 1), (red, 1), (blue, 0), (car, 1), (is, 1), (driving, 1), (too, 1), (fast, 1), (slowly, 0)}	(1, 1, 0, 1, 1, 1, 1, 1, 0)
The blue car is driving slowly	{(the, 1), (red, 0), (blue, 1), (car, 1), (is, 1), (driving, 1), (too, 0), (fast, 0), (slowly, 1)}	(1, 0, 1, 1, 1, 1, 0, 0, 1)

Note: In the term frequency lists and vectors, the numbers represent the number of times a term occurs in a document, in these examples either 1 or 0 times. For example, the first number in these two vectors indicates that the word *the* occurred once, while the second number shows that *red* occurs once in the first vector and it did not occur in the second vector.

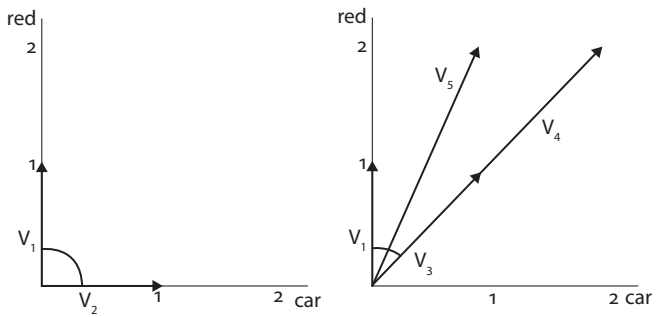

Figure 1. Visual representations of vectors from Table 4. The angle between vectors indicates the similarity between the documents they represent: the smaller the angle, the greater the similarity

Table 4. Sample 'documents' and their unweighted vectors

Document number	Document	Term frequency list	Unweighted vector
1	red	{(red, 1), (car, 0)}	$v_1 = (1, 0)$
2	car	{(red, 0), (car, 1)}	$v_2 = (0, 1)$
3	red car	{(red, 1), (car, 1)}	$v_3 = (1, 1)$
4	red car red car	{(red, 2), (car, 2)}	$v_4 = (2, 2)$
5	red red car	{(red, 2), (car, 1)}	$v_5 = (2, 1)$

Note: In the term frequency lists and vectors, the numbers represent the number of times a term occurs in a document.

There are two extreme cases. When two vectors are perpendicular to each other (that is, there is a 90-degree angle between them), the cosine is 0, indicating that the two documents represented by the vectors are completely different; when two vectors have a 0-degree angle between them, the cosine is 1, indicating that the two documents are exactly the same (ignoring the order in which the words occur). If a document is an exact copy of another document, but contains the same text twice (as in documents 3 and 4 in Table 4), the angle is 0, indicating complete overlap. However, if two documents contain the same words but with different relative frequencies, as in documents 3 and 5 in Table 4, the angle is larger than 0, indicating less than complete overlap. In contrast, the Jaccard Index in this case is 1, indicating complete overlap. This shows that the cosine similarity value, unlike the Jaccard Index, reflects the relative frequency of the words within the documents, and can distinguish between the above-mentioned retellings about cars and bikes.

With cosine similarity, the two frequency-related limitations of the Jaccard Index can be addressed: the frequency value for each word can be adjusted according to the word's frequency within iterations (term frequency) and across iterations (document frequency), so that the effect of frequent words on the similarity value is attenuated compared to less-frequent words. For the term frequency weighting the function $w(\text{tf})$ is used, if $\text{tf} > 0$.

$$w(tf) = 1+\log_2(tf), \text{ or } \text{ if } tf = 0 \text{ then } w(tf) = 0$$

where $\log_2$ is the logarithm base 2. If the word is used once, $tf = 1$ and $w(1) = 1 + 0 = 1$. If the word is used twice, $tf = 2$ and $w(2) = 1 + 1 = 2$. If $tf = 3$, then $w(3) = 1 + log_2(3) = 2.58$; and if $tf = 10$, then $w(10) = 1 + log_2(10) = 4.32$. If a word does not occur, $w(tf) = 0$. These numbers show that if we use this calculation instead of the raw frequency, the fact that a particular word was used 10 times is only weighted 4.3 times more (because $w(10) = 4.32$) than if it had been used only once ($w(1) = 1$): its effect on the similarity value is attenuated (but still more than 1). Each element (word) in the vector representing an iteration now receives $w(tf)$ as a weight, and so it becomes this:

$$v_j = (w(tf_{j,1}), w(tf_{j,2}), \ldots, w(tf_{j,n}))$$

The second frequency issue we would like to take into account is the number of times a word was used by different speakers. For this, we use the document frequency (*df*), which in this study is the number of documents (iterations) a word appears in out of all iterations by all speakers for all prompts:

$$df = \frac{\text{the number of documents the term appears in}}{\text{the total number of documents}}$$

Because we would like a word to receive a lower weight if it is more frequently used, we use the inverse of the document frequency: a high frequency results in a low value. Next, we take the log of that value to flatten the curve. In short, the document frequency weight is calculated as follows:

$$\log\left(\frac{1}{df_j}\right)$$

Next, $w(tf_j)$ is multiplied by this logarithm of the inverse document frequency of a word. The vector then becomes this:

$$v_j = \left(w(tf_{j,1})\log\left(\frac{1}{df_1}\right), w(tf_{j,2})\log\left(\frac{1}{df_2}\right), \ldots, w(tfj,_n)\log\left(\frac{1}{df_n}\right)\right) \textit{ for document } d_j$$

An example can illustrate the effect of these weight calculations (see Table 5). If a particular word is used three times in one iteration ($tf = 3$) and by 10 out of 15 speakers ($df = 10/15$), the term frequency weight is 2.58 and the logarithm of the inverse document frequency is log(15/10), so its component in the vector is not 3, but 2.58 * 0.58 = 1.51; the word receives a relatively low value because many other speakers have used it as well. In contrast, if the same word had been used by only 3 out of 15 speakers, its component would have been 2.58 * 2.32 = 6.00; this value is much higher because the word is less commonly used among other speakers and therefore more informative about a certain speaker and his or her re-use across iterations.

Table 5. Weighted values of vector components for three values of term frequency (*tf*) and three values of document frequency (*df*)

tf	*w*(*tf*)	***df* = 1/15** **log(1/*df*) = 3.91**	***df* = 3/15** **log(1/*df*) = 2.32**	***df* = 10/15** **log(1/*df*) = .058**
1	1	3.91	2.32	0.58
3	2.58	10.10	6.00	1.51
10	4.32	16.89	10.04	2.53

Note: $w(tf) = 1 + \log_2(tf)$ if $tf > 0$

To summarize, the cosine similarity method uses vectors to represent the words in the iterations, and the cosine of the angle between the vectors represents the similarity of two iterations, ranging from 0 (no similarity) to 1 (exact similarity). This method takes into account both the frequency of words within an iteration (term frequency) and the frequency of use by different speakers (document frequency), giving relatively lower values to more frequently used words. The combined weighting of term frequency and inverse document frequency is referred to as tf-idf weighting (Manning et al., 2008).

Research questions

Most studies on the effects of task repetition have examined changes in complexity, accuracy, and/or fluency and have attributed any improvements at least in part to the re-use of words and grammatical structures. But few studies have tried to determine to what extent such re-use actually occurs. These studies have used fairly intuitive measures to capture re-use at the level of individual words or exact word strings. The present study examines the extent to which immediate task repetition leads to re-use of words and grammatical structures as well as the relationship between this re-use and oral fluency. It uses a more sophisticated method (cosine similarity with tf-idf weighting) to assess re-use of both individual words and grammatical structures and to examine its relation to changes in oral fluency under conditions of immediate task repetition and increasing time pressure. The following research questions are addressed.

1. Does immediate task repetition elicit re-use at and above the word level, as represented by lexical unigrams (e.g., *the, red, car*), lexical trigrams (*the red car*), and POS trigrams (DET ADJ N)?
   To evaluate the similarity between two subsequent iterations, we first established a baseline value that represents individual speakers' similarity between two performances of the same task type but with different content, in this case the similarity between the first performances for different prompts. It is hypothesized that similarity values are higher for two performances for the same prompt than for two performances for different prompts.

2. Is there a difference in the amount of re-use at and above the word level, as represented by lexical unigrams, lexical trigrams, and POS trigrams, between conditions with constant and decreasing time?
   Since the 4/3/2 procedure was conceived as a fluency building technique that combines the supposedly beneficial effects of re-use (elicited by repetition) and

increasing time pressure, we explore whether time pressure supports or limits re-use. It is supported when speakers choose to re-use words and structures to deal with time pressure, but it is limited if they choose to change or leave out information; it is also possible that they choose a combination of strategies. Similarity values will give an indication of which strategy was chosen most.

3. Is there a difference in the amount of re-use at and above the word level between different pairs of iterations (first and second versus second and third)?
Speakers could use different strategies with different outcomes. First, it is possible that the first iteration requires most work in terms of organizing content, retrieving lemmas and building grammatical structures, which speakers can choose to refine in the second iteration, needing less refinement in the third iteration; this would result in *higher* similarity between the second and third iterations than between the first and second. Alternatively, repetition is likely to reduce working memory demands, with the largest cumulative benefit in the third iteration; if speakers use this to add content or refine lexical choices and grammatical structures, similarity would be *lower* between the second and third iterations than between the first and second. Speakers could use one or both of these strategies, thus reducing each other's effects.

4. Does re-use at and above the word level support oral fluency?
Re-use, both at the word level and above, can support fluency because it decreases attentional demands for conceptualization and formulation. It is therefore expected that similarity values correlate with oral fluency measures.

Method

Participants

Fifty-three adult learners of English as a second language were recruited in three sections of a high-intermediate speaking course in the intensive English program at a large university in the United States. Analyses were performed on data of only those participants who had signed the informed consent form and who had completed all three iterations for at least one picture story prompt. In total there were complete data sets for 39 participants (M = 26.7 years, SD = 5.1; three participants did not report their age), who came from mixed language backgrounds: Arabic (24), Chinese (5), Korean (4), Vietnamese (2), and single speakers of Farsi, Taiwanese, Thai, and Turkish. The number of complete sets per prompt was 34 for Bicycle, 30 for Tiger, and 29 for Race; 24 participants had complete data sets for all three prompts.

Materials

The prompts were three six-panel picture stories (Bicycle, Tiger, and Race), all taken from Heaton (1966), a book that has been used for several studies on second language written and oral production (N. de Jong & Vercellotti, 2016; Snellings, Van Gelderen, & De Glopper, 2004; Tavakoli & Foster, 2008). Color was added to the original black-and-white line drawings to add clarity and to highlight important elements. Four to five guiding questions were added to each prompt to stimulate speakers to talk about the same content, including the characters' emotions. All three picture stories and the guiding questions are available in the Appendix and in the IRIS digital repository of data collection instruments (Marsden, Mackey, & Plonsky, 2016). De Jong and Vercellotti (2016) argued that these three stories are similar in storyline complexity (some intentional reasoning), narrative structure (little causal reasoning), and number of elements (limited number of characters and props). They showed that these prompts elicited similar performances with respect to fluency, compositional complexity, and accuracy. The only notable difference was that Race elicited performance with a higher lexical variety than Bicycle, with fewer nouns that were used by all or most participants. Apart from this, overall, the three prompts are considered to elicit quite similar performances.

The content of the three picture stories was as follows. In the Bicycle story, a boy is pushed off the road by an angry driver and falls; after getting up, he smiles as he sees the driver by the side of the road next to his broken-down car. In Race, a tall boy tells a shorter boy he is convinced he will win the race, and after gaining a substantial lead he decides to take a rest and falls asleep, but the other runners pass him and the short boy wins. In Tiger, a hiker is warned about a tiger in the mountains; when he sits down to rest, the tiger attacks him but hits a tree and passes out before the hiker can hit him; the hiker then hits the tiger on the head with a rock, while hunters are cheering.

Procedure

All speakers retold the picture stories individually and were recorded by a computer in a language lab that was familiar to them. Prompts were displayed and retellings were recorded using a software program developed with Runtime Revolution (Shafer, 2006) run on desktop computers. For each prompt, participants were given three minutes of planning time before retelling it three times. Under the decreasing time condition, the amount of time available for the three iterations was 180, 135 and 90 seconds, respectively (shorter than the original four, three and two minutes, but in the same proportions), and speakers were encouraged to use all of the available time. Under the constant time condition, 135 seconds were available for each iteration. During planning time, the participants saw both the pictures and the questions;

during the recording, only the prompts were visible but the questions were not. A clock at the bottom of the screen showed the remaining time. Data were collected in three sessions, one for each prompt, three to four days apart. All speakers retold the picture stories in the same order: Bicycle, Tiger, Race.

Analysis

Similarity between pairs of iterations was calculated as cosine similarity with weighting for term frequency (reflecting frequency within a performance) and inverse document frequency (reflecting frequency across all performances of all speakers in this study). Similarity was based on three types of items:

- lexical unigrams: individual words in their original form, for example with inflections (counting *ride* and *rides* as two different forms);
- lexical trigrams: sequences of three words in their original form;
- part-of-speech (POS) trigrams: sequences of three parts of speech, without inflections (for example, counting *ride*, *rides, drive*, and *driving* as instances of a verb)

All trigrams ignored fillers, pauses, false starts, repetitions and corrected words; errors were not ignored. Below is an example of an AS unit (in CHAT transcription format; MacWhinney, 2000) and its corresponding unigrams and trigrams.

transcript:	and there is a boy &i <front of> [/] &hi ah <front of> [//] ah in front of car .
lexical unigrams:	and; there; is; a; boy; in; front; of; car
lexical trigrams:	_ _ and; _ and there; and there is; there is a; is a boy; a boy in; boy in front; in front of; front of car; of car _; car _ _
POS trigrams:	_ _ conj:coo; _ conj:coo pro:exist; conj:coo pro:exist v; pro:exist v det; v det n; det n prep; n prep n; prep n prep; n prep n; prep n _; n _ _

The example shows that every possible trigram was included in the analysis: moving the window across the entire utterance captures as many aspects of grammatical structure as possible. If one word is added in the next iteration (for example, *a boy* is changed to *a young boy*), two trigrams are changed and one is added, as a result of which the similarity value decreases. Special markers were added at the beginning and end of AS units (e.g., for the AS unit *He ran fast*: "_ _ he", "_ he ran" and "ran fast _", "fast _ _"). This way, structures such as *wh*-questions are captured (e.g., "_ how are") and any word at any position occurs in exactly three trigrams.

Fluency was evaluated by two measures. First, we used the proportion of time filled with speech (similar to phonation/time ratio), following the computations used by de Jong and Perfetti (2011). This measure was chosen because it reflects the total

pausing time, including both the number[2] and length of silent and filled pauses, and as such is a composite measure of temporal fluency. Second, we used mean syllable duration (the inverse of articulation rate), which has been shown to be related to linguistic skills and to increase across iterations of the 4/3/2 procedure (N. de Jong & Perfetti, 2011; N. H. de Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2013).

Results

Re-use between and within prompts

The first research question asked whether immediate task repetition of picture story retelling elicited re-use at and above the word level. As a baseline to evaluate the amount of repetition, we used the similarity between equivalent iterations of different prompts. The best comparison is between the first iterations for the prompts, because they are not influenced by content repetition and changes in available time. The cosine similarity with tf-idf weighting of each pair of iterations is shown in Table 6 for lexical unigrams, lexical trigrams, and POS trigrams, for participants with complete data sets for all three prompts.

Table 6. Cosine similarity with tf-idf weighting between the first iterations for different prompts based on lexical unigrams, lexical trigrams, and POS trigrams

		Bicycle1-Tiger1		Bicycle1-Race1		Tiger1-Race1	
		M	*SD*	*M*	*SD*	*M*	*SD*
Lexical unigrams	DT	.10	.04	.11	.06	.10	.06
	CT	.10	.07	.11	.08	.10	.07
Lexical trigrams	DT	.02	.01	.02	.01	.02	.02
	CT	.02	.02	.02	.02	.02	.02
POS trigrams	DT	.14	.09	.10	.06	.12	.08
	CT	.12	.07	.08	.04	.09	.04

Note: DT: $n = 11$; CT: $n = 13$; DT = decreasing time; CT = constant time

A multivariate GLM was run with the cosine similarity for lexical unigrams, lexical trigrams, and POS trigrams as the dependent variables, time condition (constant time, decreasing time) as between-subjects factor and prompt pair (Bicycle-Tiger,

2. Note that the number of pauses is the inverse of the mean length of fluent runs, which is another commonly used measure of temporal fluency, and is considered to reflect proceduralization (N. de Jong & Perfetti, 2011; Towell, Hawkins, & Bazergui, 1996). The proportion of time filled with speech therefore also captures proceduralization.

Bicycle-Race, Tiger-Race) as within-subjects factor. There was no overall effect of time condition, meaning that similarity values were similar whether the available time was 180 or 135 seconds, and whether the available time for subsequent iterations was decreasing or not ($F(3,21) = 0.691$, $p = .568$, partial $\eta^2 = .090$). The overall effect of prompt pair was significant ($F(6,18) = 3.474$, $p = .019$, partial $\eta^2 = .537$). Post-hoc univariate analyses showed that this effect was found only for the POS trigrams ($F(2,46) = 10.360$, $p < .001$, partial $\eta^2 = .311$), with the similarity between Bicycle and Tiger being higher than the similarity between Bicycle and Race ($p = .001$), and between Tiger and Race ($p = .045$), but no difference in similarity between Bicycle and Race, and Tiger and Race ($p = .127$). There was no interaction between time condition and prompt pair.

These data clearly show that there is some re-use of unigrams and trigrams across prompts (all lower bounds of 95% confidence intervals > 0), but this re-use was quite limited, particularly considering that some of the lexical trigrams repeated across prompts included words and bigrams at the beginning and end of AS units (e.g., *and* or *the man* occurring at the beginning of an AS unit). Not surprisingly, lexical trigram similarity was much lower than POS trigram similarity. The fairly low similarity values for lexical unigrams and trigrams can be explained by the fact that these included inflected forms, not lemmas: even one changed inflection results in an entire trigram or unigram being considered different.

Next, we compared similarity values between prompts to those within prompts (see Table 7). A multivariate GLM with repeated measures was run with the cosine similarity for lexical unigrams, lexical trigrams and POS trigrams as the dependent variables, time condition (constant time, decreasing time) as between-subjects factor, and repetition type (between prompts, within prompts) as within-subjects factor. The analysis reported here compares the similarity between the first and second iterations for the Bicycle prompt with the similarity between the first iterations for Bicycle and Tiger. The overall significance for time condition was not significant ($F(3,22) = 1.964$, $p = .149$, partial $\eta^2 = .211$). There was, however, a significant effect of repetition type, with similarity being higher between the first two iterations of Bicycle than between the first iterations of Bicycle and Tiger ($F(3,22) = 66.738$, $p < .001$, partial $\eta^2 = .901$). Post-hoc univariate analyses showed that the difference was significant for all three measures (all $ps < .001$). There was no interaction between time condition and repetition type ($F(3,22) = 1.489$, $p = .245$, partial $\eta^2 = .169$). The same pattern of results was found for the other prompts. In summary, the answer to the first research question is affirmative, in that immediate task repetition of picture story retelling elicits re-use not only at the word level but also above the word level.

Table 7. Means and SDs of the cosine similarity values with tf-idf weighting between iteration pairs for all prompts, based on lexical unigrams & trigrams, and POS trigrams

			Iterations 1&2		Iterations 1&3		Iterations 2&3	
			M	*SD*	*M*	*SD*	*M*	*SD*
Lexical unigrams	Bicycle	DT	.47	.10	.41	.11	.58	.11
		CT	.51	.16	.48	.14	.63	.12
	Tiger	DT	.52	.10	.47	.09	.57	.07
		CT	.57	.12	.54	.09	.61	.11
	Race	DT	.51	.11	.49	.10	.61	.13
		CT	.58	.10	.56	.08	.66	.08
Lexical trigrams	Bicycle	DT	.14	.04	.11	.04	.19	.07
		CT	.18	.07	.18	.07	.25	.10
	Tiger	DT	.16	.06	.15	.07	.21	.05
		CT	.21	.09	.19	.07	.25	.10
	Race	DT	.17	.09	.16	.06	.23	.11
		CT	.21	.07	.20	.06	.27	.09
POS trigrams	Bicycle	DT	.24	.08	.20	.09	.30	.11
		CT	.28	.12	.26	.11	.34	.09
	Tiger	DT	.30	.10	.22	.06	.30	.04
		CT	.29	.11	.29	.09	.34	.11
	Race	DT	.29	.12	.28	.10	.34	.14
		CT	.30	.09	.28	.07	.35	.10

Note: DT: $n = 11$; CT: $n = 13$; DT = decreasing time; CT = constant time

Re-use under constant and decreasing time and by iteration

The second and third research question asked whether the amount of re-use at and above the word level varied by time condition and by iteration pair. A multivariate GLM analysis was run with the cosine similarity for lexical unigrams, lexical trigrams, and POS trigrams as the dependent variables, time condition (constant time, decreasing time) as a between subjects factor, and prompt (Bicycle, Tiger, Race) and iteration pair (first-second, second-third) as within subjects factors. Because similarity between the first and third iteration is not independent of similarity between the first and second and between the second and third, we only compared the similarity between the first and second iteration to the similarity between the second and third.

No statistically significant difference was found between the two time conditions ($F(3,20) = 1.552$, $p = .232$, partial $\eta^2 = .189$). There was, however, a significant effect of iteration pair ($F(3,20) = 11.756$, $p < .001$, partial $\eta^2 = .638$). Post-hoc univariate analyses

showed that this effect was found for all three measures (all *ps* < .001). There were no interactions between time condition, prompt and/or iteration pair. Taken together, these results show that a decrease in available time does not affect re-use at the word level or above, but more re-use was elicited by the third than by the second iteration. Finally, an overall main effect of prompt was found ($F(6,17) = 6.624$, $p = .001$, partial $\eta^2 = .700$), which post-hoc univariate analyses showed was driven by lexical unigrams only ($F(2,44) = 13.302$, $p < .001$, partial $\eta^2 = .377$). Pairwise comparisons showed that similarity for lexical unigrams was lower for Tiger than for Bicycle and Race (both *ps* < .001), while there was no difference between Bicycle and Race ($p = .348$).

Correlations between similarity and fluency

To answer the fourth research question, which asked whether re-use at and above the word level supports oral fluency, we examined the Pearson correlations between

Table 8. Means and standard deviations for proportion of time filled with speech and mean syllable duration

		Iteration 1		Iteration 2		Iteration 3	
		M	*SD*	*M*	*SD*	*M*	*SD*
Proportion of time filled with speech	Bicycle (*n* = 34)	.57	.10	.59	.10	.61	.10
	Tiger (*n* = 30)	.57	.08	.60	.08	.64	.08
	Race (*n* = 29)	.60	.07	.62	.08	.65	.07
Mean syllable duration	Bicycle (*n* = 34)	.32	.04	.32	.04	.30	.04
	Tiger (*n* = 30)	.32	.03	.30	.04	.29	.04
	Race (*n* = 29)	.33	.05	.31	.04	.30	.04

Table 9. Partial pearson correlations between cosine similarity values (iterations 1 and 3) and two measures of fluency on iteration 3, controlling for the same fluency measures on iteration 1

		Lexical unigrams		Lexical trigrams		POS trigrams	
		r	*p*	*r*	*p*	*r*	*p*
Proportion of time filled with speech	Bicycle (*n* = 34)	.134	.458	.266	.135	.416*	.016
	Tiger (*n* = 30)	−.493*	.007	−.490*	.007	−.338†	.073
	Race (*n* = 29)	−.133	.501	−.218	.265	.148	.452
Mean syllable duration	Bicycle (*n* = 34)	−.084	.644	−.051	.777	.008	.964
	Tiger (*n* = 30)	.404*	.030	.273	.151	.456*	.013
	Race (*n* = 29)	.147	.454	.245	.209	.272	.161

Note: † $p < .10$; * $p < .05$.

cosine similarity values on the one hand, and on the other hand two fluency measures at the third iteration for each prompt: the proportion of time filled with speech – a composite measure of fluency – and mean syllable duration (descriptive statistics in Table 8). To control for individual differences in fluency that are unrelated to the task but rather a characteristic of the speaker, fluency at the first iteration of the same prompt (either the proportion of time filled with speech or syllable duration) was controlled for. For the proportion of time filled with speech, as Table 9 shows, there was a medium correlation for Bicycle for the similarity value based on POS trigrams, but not for lexical trigrams or unigrams. Unexpectedly, there were medium but negative correlations between this fluency measure and similarity for Tiger, and no correlations for Race. For mean syllable duration similar medium correlations were found, again only for the Tiger prompt. Note that correlations are positive here because a higher value for mean syllable duration indicates lower fluency. In summary, similarity did not correlate with fluency as expected, in that speakers whose iterations were more similar tended to speak less fluently, but only for the Tiger prompt.

Discussion and conclusion

This study has shown that re-use of words and grammatical structures between repeated picture story retellings is higher than between retellings of two different stories. This is the case not only at the level of individual words (unigrams), as de Jong and Perfetti (2011) found, but also at a higher level (lexical trigrams and POS trigrams). This finding is relevant for second language acquisition and teaching, because if repeated use of language is more likely when tasks are repeated with the same content than with new content, specific task repetition is more likely to stimulate proceduralization and automatization of linguistic knowledge.

In classrooms, task repetition is often combined with increasing time pressure in the 4/3/2 activity. Our analyses suggest that decreasing the amount of time available for subsequent iterations does not affect re-use at the word level or above (trigrams). These findings contrast with Thai and Boers's (2016) and Boers's (2014) findings, where similarity was higher in the decreasing time condition than in the constant time condition. One explanation for this difference can be found in students' familiarity with the task content. In Boers's and Thai and Boers's studies, participants spoke about familiar topics, for example, their family, home town, memorable events or their favorite movie, which means that not only the message content but also the words and grammatical structures may have been very familiar and easily retrieved. This is supported by Boers's finding that some speakers repeated n-grams of 30 words and more; this far exceeds working memory and suggests some stretches of speech had been memorized before the study. In the present study, on the other hand, participants

retold a picture story that was new to them, which required them to not only generate and retrieve new content, words, and grammatical structures, but also encouraged them to refine these as they repeated their retellings.

Although the similarity values showed no effect of time pressure, there seem to have been differences with respect to re-use in terms of the addition and removal of information. This is suggested by the number of words used in each iteration: speakers in the constant time condition added on average about 25 words (15%) and 5 clauses to their final iteration for a particular prompt compared to the first. But this effect was not as large as in the decreasing time condition, where speakers, for all three prompts, cut about 40 to 50 words (25%[3]) and 9 clauses from their final iteration compared to the first. Inspection of the transcripts shows that under time pressure, particularly in the final iteration, some speakers left out evaluative or summarizing comments, such as *and from this story we should learn that bothering anybody will bother our self*. Conversely, speakers under the constant time condition often added information about the plot (e.g., *then the man drive away*). This pattern is similar to Boers's observation, although in his study speakers cut the number of words by 42% from the first to the third iteration. This may be another explanation for the larger effects of time pressure on similarity in his study compared to the current study. Further qualitative analysis is needed to examine exactly what speakers in the two time conditions added or left out.

This study also showed that similarity was higher between the second and third iteration than between the first and second. This difference was found despite the three-minute planning time given before the first retelling, in which speakers could have rehearsed the task. This finding supports the argumentation by Bygate and Samuda (2005) and Skehan (2014) that there is a difference between pre-task planning and the first iteration in task repetition. While many speakers during pre-task planning may perform lexical searches, retrieve grammatical structures, or even rehearse (cf. Ortega, 2005), the actual performance of a task forces all speakers to generate words and grammatical structures for all or most of the content to be expressed. These can be revised in the second iteration but may need less revision in the third iteration. The difference between the similarity of the two pairs of iterations was not uniform across prompts, however: it was smaller for Tiger, the second prompt. Furthermore, similarity was lowest between the first and second iteration for the first prompt, and highest between the second and third iteration of the last prompt. It therefore seems

3. Although a difference in text length may affect the reliability of similarity measures, this becomes important only with large differences. In our study, the largest differences in text length (number of tokens) were found between the first and final iteration in the decreasing time condition for Bicycle and Race, where the shorter iteration had 74% of the longer iteration. We deem this an acceptable length difference.

possible that, for the first prompt, speakers were still unsure how to repeat a task most efficiently, while on the last iteration for the last prompt they had developed some strategies.

The final analysis showed that the relationship between similarity and fluency is variable.[4] Partial correlations between similarity and time filled with speech were positive for POS trigrams for the first prompt, but negative for unigrams and trigrams for the second prompt, and absent for the third prompt. The positive correlations for the first prompt, Bicycle, confirm the expectation that re-use supports fluency, but only at the most abstract level and for one fluency measure. The most likely explanation is that speakers re-used grammatical structures as a strategy to decrease attentional demands for conceptualization and formulation and to increase their fluency. However, on the second prompt, Tiger, similarity was related to less fluent speech; that is, similarity values not only correlated negatively with the proportion of time filled with speech but also positively with mean syllable duration. This suggests that those speakers who repeated the most words and structures also struggled most with fluency; for them, re-use may have been a strategy to deal with the high demands of the task. This is consistent with Boers's observation that some participants reported struggling to decide what information to leave out when time decreased or had difficulty remembering what they had said previously. The absence of a correlation for the third prompt suggests that different speakers used different strategies. The mixed effects can be interpreted as highlighting that speakers have options for how to use their attentional demands under task repetition and time pressure conditions. Their choice may depend on the prompt or on their experience with the task type or task conditions: having experienced increasing time pressure once, on a second or third occasion, speakers may deal with it differently. Future research will need to show how and why speakers decide to use their attentional resources to increase complexity, accuracy, and/or fluency.

Taken together, these analyses show that repetition of the task-as-workplan can result in considerable repetition of the task-as-process. Re-use, whether verbatim in the lexical trigrams, or abstract in the POS trigrams, may stimulate the development of oral fluency. Whether accuracy and complexity are stimulated has not been addressed in this study. Thai and Boers (2016) showed that many errors were repeated across iterations. If accuracy and complexity are the goal, task repetition may therefore need to be supplemented with a form-focused stage between iterations (cf. Hawkes, 2012; Lyster & Sato, 2013).

4. Correlation analyses with other fluency measures, such as the number of repeated or reformulated words, did not yield any significant results. These measures are not reported due to space limitations.

In terms of research methodology, the present study shows that insights from the field of information retrieval can be applied to task repetition data to estimate the amount of re-use at the word level and above.[5] With relatively straightforward formulas, large amounts of data can be analyzed without time-consuming manual annotations of re-used strings of words. With tf-idf weighting, the similarity measure takes into account the frequency of words, so that function words and highly frequent content words do not unduly inflate similarity estimates. In this paper, we have reported analyses at the lexical and part-of-speech levels, which represent two ends of the spectrum. Our analysis at the lemma level, not reported here, yielded very similar results: higher similarity between the second and third iterations than between the first and second, and a trend indicating higher trigram similarity for the constant time condition compared to the decreasing time condition.

One potential limitation to the measures used here is the insensitivity to the location of the words or trigrams. If a speaker on the first iteration uses *the man was angry and annoyed* and on the second iteration *the man was annoyed and angry*, these two utterances are considered dissimilar in terms of lexical trigrams, although the grammatical structure is the same. Adding lexical unigrams and POS trigrams somewhat compensated for this problem, as the two utterances have the same lexical unigrams and the same POS trigram, *adj conj adj*. Another limitation is that lexical semantics is not taken into account. If a speaker on the first iteration uses the word *bicycle* but he/she uses *bike* on the second iteration, these two words are considered different, although they are more similar than *bicycle* and *vehicle*, for instance. Although this is a concern, there is no straightforward solution, if only because the distinction between similar and dissimilar meanings is not clear-cut.

In conclusion, the present study shows that immediate repetition of task type and content results in considerable re-use, which may stimulate the proceduralization and automatization of linguistic knowledge and therefore second language development. Time pressure, however, does not seem to affect re-use, at least when speakers speak about a topic that is novel to them, in this case a picture story. Re-use was found not only at the word level but also above, as evidenced by the lexical trigrams and POS trigrams, and seemed to be used as a strategy by speakers struggling with fluency. On this basis, it seems that immediate specific task repetition, whether under increasing time pressure or not, can support fluency development through re-use of individual words and grammatical structures.

5. Although lexical trigrams and POS trigrams may be used effectively with analytic languages such as English, their use is much less effective with agglutinative languages. To assess similarity at the morphological level, whether in analytic or agglutinative languages, trigrams of phonemes can be used, but this was outside the scope of the present paper.

Acknowledgements

We would like to thank Martin Bygate and an anonymous reviewer for their insightful comments on this article. We are also grateful to Laura Halderman, Jon-Michel Seman, and Mary Lou Vercelotti for the dedication and effort they put into collecting and annotating the data. This work was supported in part by the Pittsburgh Science of Learning Center, which was funded by the National Science Foundation award number SBE-0836012.

References

Ahmadian, M. J., & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research*, 15, 35–59. https://doi.org/10.1177/1362168810383329

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard & S. Oksefjell (Eds.), *Out of corpora: Studies in honour of Stig Johansson* (pp. 181–189). Amsterdam: Rodopi.

Biber, D., Conrad, S., & Cortes, V. (2004). *If you look at…*: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25, 371–405. https://doi.org/10.1093/applin/25.3.371

Boers, F. (2014). A reappraisal of the 4/3/2 activity. *RELC Journal*, 45, 221–235. https://doi.org/10.1177/0033688214546964

Breen, M. 1989. The evaluation cycle for language learning tasks. In R.K. Johnson (Ed.), *The second language curriculum* (pp. 187–206). Cambridge: Cambridge University Press.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 23–48). Harlow: Pearson Longman.

Bygate, M., & Samuda, V. (2005). Integrative planning through the use of task-repetition. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 37–74). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.11.05byg

Chen, Y-H., & Baker, P. (2016). Investigating criterial discourse features across second language development: Lexical bundles in rated learner rssays, CEFR B1, B2 and C1. *Applied Linguistics*, 37, 849–880.

De Cock, S. (2004). Preferred sequences of words in NS and NNS speech. *Belgian Journal of English Language and Literatures*, 2, 225–246.

de Jong, N. (2012). Does time pressure help or hinder oral fluency? In N. De Jong, K. Juffermans, M. Keijzer, & L. Rasier (Eds.), *Papers of the Anéla 2012 Applied Linguistics Conference* (pp. 43–52). Delft: Eburon.

de Jong, N., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, 61, 533–568. https://doi.org/10.1111/j.1467-9922.2010.00620.x

de Jong, N., & Vercellotti, M. L. (2016). Similar prompts may not be similar in the performance they elicit: Examining fluency, complexity, accuracy, and lexis in narratives from five picture prompts. *Language Teaching Research*, 20, 387–404.

de Jong, N. H., Steinel, M., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, 34, 893–916. https://doi.org/10.1017/S0142716412000069

DeKeyser, R. M. (2009). Cognitive-psychological processes in second language learning. In M. Long & C. Doughty (Eds.), *The handbook of second language teaching* (pp. 119–138). Oxford: Blackwell. https://doi.org/10.1002/9781444315783.ch8

Ellis, R. (2003). Tasks in SLA and language pedagogy. In R. Ellis (Ed.), *Task-based language learning and teaching* (pp. 1–35). Oxford: Oxford University Press.

Francis, W. S., Corral, N. I., Jones, M. L., & Sáenz, S. P. (2008). Decomposition of Repetition Priming Components in Picture Naming. *Journal of Experimental Psychology: General*, 137, 566–590. https://doi.org/10.1037/0096-3445.137.3.566

Gass, S., Mackey, A., Alvarez-Torres, M. J., & Fernández-García, M. (1999). The effects of task repetition on linguistic output. *Language Learning*, 49, 549–581. https://doi.org/10.1111/0023-8333.00102

Hawkes, M. L. (2012). Using task repetition to direct learner attention and focus on form. *ELT Journal*, 66, 327–336. https://doi.org/10.1093/elt/ccr059

Heaton, J. B. (1966). *Composition Through Pictures*. Essex: Longman.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27, 4–21. https://doi.org/10.1016/j.esp.2007.06.001

Jurafsky, D., & Martin, J. H. (2008). *Speech and language processing: An introduction to natural language processing, computational linguistics and speech recognition* (2nd ed.). Upper Saddle River, NJ: Pearson/Prentice Hall.

Kim, Y., & Tracy-Ventura, N. (2013). The role of task repetition in L2 performance development: What needs to be repeated during task-based interaction? *System*, 41, 829–840. https://doi.org/10.1016/j.system.2013.08.005

Lauttamus, T., Nerbonne, J., & Wiersma, W. (2007). Detecting syntactic contamination in emigrants: The English of Finnish Australians. *SKY Journal of Linguistics*, 20, 273–307.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

Levelt, W. J. M. (1999). Producing spoken language: A blueprint of the speaker. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 83–122). Oxford: Oxford University Press.

Lyster, R., & Sato, M. (2013). Skill Acquisition Theory and the role of practice in L2 development. In M. del P. García Mayo, M. J. Gutierrez Mangado, & M. Martinez Adrian (Eds.), *Contemporary approaches to second language acquisition* (pp. 71–91). Amsterdam: John Benjamins. https://doi.org/10.1075/aals.9.07ch4

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511809071

Marsden, E., Mackey, A., & Plonsky, L. (2016). The IRIS Repository: Advancing research practice and methodology. In A. Mackey & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS repository of instruments for research into second languages* (pp. 1–21). New York NY: Routledge.

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, 33, 299–320. https://doi.org/10.1093/applin/ams010

McCarthy, P. M., Hall, C., Duran, N. D., Doiuchi, M., Fujiwara, Y., Duncan, B., & McNamara, D. S. (2009). Analyzing journal abstracts written by Japanese, American, and British scientists using Coh-Metrix and the Gramulator. *The ESPecialist*, 30, 141–173.

McDonough, K., & Trofimovich, P. (2009). *Using priming methods in second language research*. New York, NY: Routledge.

Nation, P. (1989). Improving speaking fluency. *System*, 17, 377–384. https://doi.org/10.1016/0346-251X(89)90010-9

Ortega, L. (2005). What do learners plan? Learner-driven attention to form during pre-task planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 77–109). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.11.07ort

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, 30, 590–601. https://doi.org/10.1093/applin/amp045

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, 32, 130–149. https://doi.org/10.1017/S0267190512000098

Sample, E., & Michel, M. (2014). An exploratory study into trade-off effects of complexity, accuracy, and fluency on young learners' oral task repetition. *TESL Canada Journal*, 31, 23–46.

Shafer, D. (2006). *Revolution: Software at the speed of thought*. Monterey, CA: Shafer Media.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31, 487–512. https://doi.org/10.1093/applin/amp058

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30, 510–532. https://doi.org/10.1093/applin/amp047

Skehan, P. (2014). Limited attentional capacity, second language performance, and task-based pedagogy. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 211). Amsterdam: John Benjamins. https://doi.org/10.1075/tblt.5.08ske

Snellings, P., Van Gelderen, A., & De Glopper, K. (2004). The effect of enhanced lexical retrieval on second language writing: A classroom experiment. *Applied Psycholinguistics*, 25, 175–200. https://doi.org/10.1017/S0142716404001092

Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, 58, 439–473. https://doi.org/10.1111/j.1467-9922.2008.00446.x

Thai, C., & Boers, F. (2016). Repeating a monologue under increasing time pressure: Effects on fluency, complexity, and accuracy. *TESOL Quarterly*, 50, 369–393.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, 17, 84–119. https://doi.org/10.1093/applin/17.1.84

Appendix

Picture story prompts and guiding questions for Bicycle, Tiger and Race.

Bicycle
In picture 1, what is the driver of the car doing?
Why can't the car pass the boy on the bicycle?
In picture 6, what is the boy doing?
How do the people feel?

Tiger
In picture 1, what can you see on the hotel wall?
What is the man on the right trying to do? Why?
In picture 4, what did the man try to do?
In picture 5, what has the tiger done?
In picture 6, what are the three men doing?

Race
In picture 1, what might the tall runner be saying?
Why does the tall runner decide to take a nap?
In picture 5, what is the tall runner thinking?
Who wins the race?
How do the boys feel in the end?

CHAPTER 3

The effects of task repetition and task complexity on L2 lexicon use

YouJin Kim,[1] Scott Crossley,[1] YeonJoo Jung,[1] Kristopher Kyle[2] and Sanghee Kang[1]
[1]Georgia State University / [2]University of Hawai'i at Mãnoa

To date, a growing number of studies in instructed second language acquisition (SLA) have investigated the effects of task repetition on L2 performance (e.g., complexity, accuracy, fluency) as well as on the occurrence of interaction-driven learning opportunities (e.g., Bygate, 2001; Fukuta, 2016; Kim & Tracy-Ventura, 2013). However, studies have not investigated how repeating different aspects of tasks (e.g., content or procedure) during collaborative tasks affects second language lexical use. The current study aims to fill that gap by comparing the impact of procedural repetition and task repetition on the lexical use of paired language learners as they repeat collaborative tasks. Furthermore, how task complexity mediates task repetition effects was also investigated. Four intact junior high school English classes (n = 73) were randomly assigned to one of the following groups: (1) simple task with exact task repetition, (2) simple task with procedural repetition, (3) complex task with exact task repetition, and (4) complex task with procedural repetition. All groups participated in three collaborative tasks over three days. The task repetition groups carried out the same information-exchange task with the same content three times, while the procedural repetition groups followed the same task procedure but had different content for the three sessions. Learners' interaction data during collaborative tasks were transcribed and analyzed for lexical use using a natural language processing tool. Lexical use was operationalized by word familiarity, age of acquisition (AoA), and word frequency (Kyle & Crossley, 2015). The findings indicated that as learners repeated tasks, they used words that are less familiar and have a higher age of acquisition. Task complexity was found to be a mediating factor for the relationship between task repetition and vocabulary use.

The effects of task repetition and task complexity on L2 lexicon use

Over the last two decades, tasks have been used in various empirical and pedagogical domains in the field of second language acquisition (SLA) (Bygate, 2016; Long, 2015;

DOI 10.1075/tblt.11.03kim

Van den Branden, Bygate, & Norris, 2009). The growing interest in this area has largely been due to the notion that tasks provide contexts for meaningful second language (L2) learning and use. Accordingly, tasks have become central to both L2 pedagogy and research, and have connected these two domains, especially in the instructed SLA literature (Mackey, 2007). Various task design and implementation variables have been suggested as factors which can significantly affect the amount of interaction and the quality of task-based language performance (see Kim, 2015 for a review). Although most research into tasks has concentrated on one-time task performance, the concept of task repetition introduces a time-wise dynamic into studies, which is especially relevant to pedagogical contexts. In a research synthesis paper which explored task-based learner performance research, Plonsky and Kim (2016) reported that the majority of previous task-based studies have focused on interactional features as well as complexity, accuracy, and fluency (CAF) of L2 production. In terms of linguistic foci, although interactional studies have examined students' attention to vocabulary using the analysis of language related episodes (LREs, dialogue where learners discuss language issues and correct errors during communication, Swain & Lapkin, 1998), very little research has focused on the nature of students' vocabulary production during collaborative tasks.

The purpose of the current study is to examine the role of repeating different characteristics of tasks on students' vocabulary production and to explore whether task complexity mediates the effects of task repetition on the use of sophisticated lexical items. In the current chapter we use the term "task repetition" as a general term to indicate any type of task repetition, and use different terms to refer to more specific task repetition conditions (i.e., exact task repetition: repeating both content and procedure of tasks; procedural repetition: repeating procedure with different content; and content repetition: repeating the content with different procedures).

Background

Task repetition and L2 oral production during monologic tasks

Bygate (1996, 2001) examined the psychological benefits of task repetition in L2 learning based on Levelt's (1989) speech production model. According to Levelt's model, the process of speech production is accomplished through three related phases of processing: conceptualization, formulation, and articulation. In the conceptualization phase, intentions and relevant information to be conveyed are generated in the form of a pre-planned message, or preverbal message. The preverbal message, which is propositional and conceptual in nature, is then transformed into linguistic (lexical, grammatical, and morpho-phonological) structures in the formulation stage. In the articulation phase, the linguistic structures are converted to actual speech. While each

of these phases occurs independently, resources are passed on from the preceding to the following phase. Bygate (2001) adopted this model to address the effects of task repetition on L2 development, suggesting that when carrying out oral tasks, learners are primarily concerned with content generation. More specifically, although learners may encounter challenges at the three different processing phases involved in oral production, task repetition could help familiarize them with the message content, thereby allowing them to more carefully select and monitor language forms on subsequent iterations. By repeating a similar task, learners can build on what they have previously done and "buy time" for cognitive processes involved in message generation (Ahmadian, 2013). Task repetition may, then, help learners enhance their speech performance in terms of complexity, accuracy, and fluency.

To date, empirical studies have documented some of the effects of task repetition on the development of L2 oral performance. There have been various ways to operationalize task repetition: (1) repeating the exact task (i.e., exact task repetition), (2) repeating the procedure of tasks with different content (i.e., procedural repetition), and (3) repeating the content of tasks with different procedures (i.e., content repetition). For example, Gass, Mackey, Alvarez-Torres, and Fernández-García (1999) explored the effects of task repetition on L2 Spanish learners in terms of the development of overall L2 proficiency, morpho-syntax, and lexical sophistication. Results demonstrated that task repetition was beneficial for the development of overall proficiency, selected morphosyntactic forms (*ser* and *estar; to be* in English), and lexical sophistication (e.g., lexical originality, lexical density, lexical variation, lexical frequency profiles). With regard to the use of vocabulary, the results showed that less common words were used more as the learners repeated the tasks. However, the learning effects were not carried over to the new context. In a more tightly controlled research setting, Bygate (2001) compared task performance of two task repetition groups who completed both narrative and interview tasks and a control group. The findings suggest that repeating the exact same task was beneficial for improving complexity (number of words per T-unit) and fluency (number of pauses) but not accuracy (number of errors per T-unit). However, practicing the same task type did not affect learners' performance in new versions of that same task type.

In a small-scale study, Hawkes (2012) also focused on how task repetition impacts the accuracy of learners' oral production. He created a form-focus stage, whereby learners carried out consciousness-raising activities after completing the main task, practicing target structures and vocabulary. This post-task activity was designed to help learners notice useful language for more successful task completion in the future. Learners were then asked to repeat the main task in exactly the same way with the same interlocutor. Results demonstrated that task repetition could be used as a technique to draw learners' attention to form, showing that learners paid more attention to accurate production.

Recently, Fukuta (2016) used methodological triangulation (i.e., a combination of CAF analysis and a retrospective interview [a stimulated recall protocol]) to examine whether task repetition would affect attention orientation. Twenty-eight Japanese advanced English as a Foreign Language (EFL) learners were randomly assigned to either the same task group (i.e., exacept task repetition) or the same task type group (i.e., procedural repetition). The participants carried out a narrative task with the researcher with a week's interval between the first and the second performance. Fukuta found that accuracy (the percentage of error-free AS-units) and lexical variety (Guiraud Index [GI], which calculates the number of word types over the square root of the number of word tokens), had significantly improved in the learners' second performance of the exact same task. Analysis of the interview data revealed that whereas the procedural repetition group did not exhibit apparent shifts of attention orientation, the task repetition group focused more attention on syntax and less on lexical choice at the second performance. This finding may indicate that when learners carry out the same task twice, they tend to reuse the previous lexical encoding process via memory retrieval and allocate their attentional resources to formal aspects of processing (i.e., syntactic encoding).

In sum, previous studies which implemented monologic tasks suggest positive task repetition effects; however, the effects of task repetition and its impact on different dimensions of linguistic performance (i.e., CAF) have varied across studies. The discrepancy in previous findings might be due to how task repetition is operationalized and the interval between task performance and the number of repetitions.

Repeating different characteristics of collaborative tasks

While the aforementioned studies focused on the impact of task repetition on monologic oral production (e.g., Gass et al., 1999; Bygate, 2001), an increasing number of task repetition studies have implemented collaborative tasks, which is the focus of the current study. Some studies examined the role of different characteristics of task repetition (i.e., task repetition, procedural repetition, and content repetition) on various aspects of L2 learning. For instance, Lynch and Maclean (2000) examined the relationship between procedural repetition and task performance (e.g., accuracy and fluency) by using a task called "poster carousel" during which learners at different proficiency levels made a poster based on a research article and visited each other's posters to ask questions. The carousel is not exact repetition of the same task but more of *recycling*, or *retrial* (italics in original) because the content and emphasis may vary depending on the visitor's questions, whereas the learners' communication goal remains the same across the iterations of the task. Findings demonstrated that both higher and lower proficiency learners benefited from immediate repetition of tasks, exhibiting increased accuracy and fluency of oral production. This finding supports

the notion that learners can exploit task recycling for gaining familiarity with the content and task demands as well as with the process of meaning formulation, allowing them to devote more attention to language (e.g., Lynch, 2007).

Recent classroom-based studies have implemented collaborative tasks in EFL classrooms, asking students to repeat different aspects of tasks, and then focusing on learners' attention to target grammatical features in addition to oral production (e.g., fluency, accuracy). Patanasorn (2010) examined how procedural, content, and task repetition might enhance learners' accuracy (e.g., the percentage of error-free AS-units, percentage of target-like simple past use) and fluency (the number of syllables per minutes) using collaborative film retell tasks. Results showed that only the procedural repetition group developed their accurate use of simple past while the other two groups showed no improvement. On the other hand, the content repetition group improved in terms of global fluency (number of syllables per minute). However, no significant changes were observed in the task repetition group. Overall, these findings suggest that different characteristics of task repetition may be beneficial for different dimensions of oral performance.

While Patanasorn (2010) focused on learning outcomes of different aspects of task repetition, Kim (2013) examined the occurrence of learning opportunities during collaborative tasks under two task repetition conditions (i.e., exact task repetition vs. procedural repetition). Kim compared these two task repetition conditions in terms of their effects on Korean EFL adolescent learners' attention to linguistic form during task completion. The exact task repetition group repeated the same information-exchange task three times while the procedural repetition group carried out three tasks that contained different content, following the same procedure over a one-week period. Learners' attention to linguistic form was analyzed in terms of their production of LREs. Results revealed that the procedural repetition group produced significantly more LREs than the task repetition group. Furthermore, while both learners and the teacher reported that they found the tasks helpful in learning English, concerns were expressed over repeating the exact same task due to potential decrease in learners' interest.

As a follow-up study, Kim and Tracy-Ventura (2013) investigated the impact of different characteristics of task repetition on the development of L2 task performance (i.e., CAF) by employing a pretest/posttest design. A total of 36 Korean EFL adolescent students from two intact classes participated in the study and each class was assigned to one of two groups: exact task repetition or procedural repetition. Results indicated that both types of task repetition helped to improve learners' accurate use of target linguistic features (simple past morphology) while the procedural repetition group also exhibited more marked development in syntactic complexity as measured by complexity by subordination (mean number of clauses per AS-unit) and the proportion of complex AS-units. However, neither group showed significant change in speech rate after completing the collaborative treatment tasks.

Overall, there has been a clear methodological pattern in task repetition research. While studies which implemented monologic oral tasks focused on learners' linguistic performance during tasks, those which focused on dialogic tasks investigated learning outcomes of the task performance via a pre-posttest. This shows a clear gap in the research regarding the characteristics of linguistic performance during dialogic oral tasks.

Beyond task repetition effects: Mediating variables

Researchers have recently begun addressing the potential interactions between task repetition and other task and learner variables (e.g., individual differences in working memory capacity [WMC], task planning, task complexity). For example, one of the individual differences that has been examined is WMC. Ahmadian (2013) studied the interaction between WMC and task repetition and its impact on CAF in L2 oral production. Forty-two Iranian EFL learners conducted an oral narrative task twice with one-week interval. WMC was measured through a listening span task administered in the participants' L1. Results indicated that while the relationship between WMC and learners' oral production was not significant on the first task performance, participants with larger WMC produced more fluent and accurate speech on the second performance. Based on this finding, Ahmadian suggested that learners with greater WMC are likely to be more successful in drawing on the task-based implementation factors employed in the classroom and that WMC training would be helpful for language learners.

In terms of task implementation factors, Ahmadian and Tavakoli (2011) examined the effects of simultaneous use of careful online planning and task repetition on CAF in Iranian EFL learners' oral production. Four experimental conditions were created by using two types of online planning (careful and pressured) with task repetition involving story telling tasks: careful online planning without exact task repetition, careful online planning with exact task repetition, pressured online planning with exact task repetition, and pressured online planning without exact task repetition. Results indicated that learners engaging in careful online planning with exact task repetition significantly developed with regard to all three dimensions of task performance (i.e., complexity, accuracy, fluency). Specifically, as careful online planning allowed learners to spend as much time as they needed to complete the task, contrary to pressured online planning, they reasoned that the learners could pay more attention to accurate and complex language production in a fluent manner.

Recently, Kim and Payant (2014) examined the interaction effects of task complexity and task repetition on the occurrence of interaction-driven learning opportunities when mediated by different characteristics of task repetition. Learning opportunities were operationalized by the occurrence of LREs. The findings showed that

although task complexity did not significantly impact the number of LREs, exact task repetition played a significant role in that the procedural repetition condition promoted a greater amount of LREs (i.e., attention to linguistic features during collaborative tasks) than the exact task repetition condition regardless of task complexity. The resolution of LREs was not influenced by task complexity and task repetition as approximately 80% of LREs were resolved successfully regardless of the conditions.

As reviewed thus far, a substantial body of instructed SLA research has been conducted to understand the effects of task repetition on L2 performance (e.g., CAF) (e.g., Bygate, 2001) as well as on the occurrence of interaction-driven learning opportunities (Kim & Payant, 2014). Research has also shed light on the interactive effects of task repetition and various mediating factors on learners' task performance (e.g., Ahmadian & Tavakoli, 2011). However, while the vast majority of the previous research has focused on aspects such as complexity, accuracy, fluency, overall proficiency, or morphosyntactic development, studies have not investigated how different task repetition conditions (content or procedural) during collaborative tasks might affect L2 learners' lexical production. Task repetition studies which used monologic tasks have tended to focus more on CAF measures, whereas those which implemented collaborative tasks have focused on interactional features such as LREs. In the current study, however, we focus on three L2 lexical sophistication variables related to word exposure: word frequency, word familiarity, and age of acquisition. All of these variables are related to word exposure and all have been shown to be strong predictors of L2 vocabulary acquisition and development. For instance, word familiarity, frequency, and age of acquisition scores correlate with human judgments of lexical proficiency (Crossley, Salsbury, McNamara, & Jarvis, 2011a, 2011b; Kyle & Crossley, 2015), holistic writing proficiency scores (Guo, Crossley, & McNamara, 2013; Kyle & Crossley, 2016), and holistic scores of speaking proficiency (Crossley & McNamara, 2013; Kyle & Crossley, 2015). These studies showed that written and spoken L2 texts that include words with lower familiarity scores or higher age of acquisition scores tend to earn higher proficiency/quality scores.

Additionally, previous studies have suggested a potential mediating role for other task and learner factors; however, thus far, the role of task design factors such as task complexity is not clear. This chapter aims to fill that gap by comparing the impact of procedural repetition and task repetition on the lexical production of paired language learners as they repeat collaborative tasks. It further examines whether task repetition effects are similar for tasks with different complexity levels. Hence, as with other studies in this volume, this investigation goes beyond exploring a single overarching task repetition effect, and is thus guided by the following research question: How do different task repetition conditions (i.e., procedural repetition vs. exact task repetition) and task complexity affect learners' use of sophisticated lexical items in terms of word familiarity, word age of acquisition, and word frequency?

Method

Participants

Participants were 73 female Korean middle school students enrolled in the same private, all-girls school in South Korea. They ranged in age from 12 to 15 years, with a mean age of 13.57 years ($SD = 0.71$). All participants had taken at least four years of required English classes in elementary school. The average amount of previous English instruction was 6.08 years ($SD = 2.79$). The school offered a four-hour mandatory English course per week. Out of four hours, their regular Korean English teacher taught three hours, and a native English teacher was in charge of one hour of instruction per week. While the Korean teacher focused more on teaching reading, grammar, and vocabulary, the students had opportunities to practice speaking and listening skills with the native English speaker teacher. In addition to their regular English classes at school, most students reported taking extra English lessons outside the school ($M = 5$ hours per week; $SD = 3.21$), with an average of two hours spent practicing oral communication skills in English.

Materials

Three communicative tasks were designed primarily to relate to students' interests. Because the real-world needs for EFL learners tend to be limited, topics for the tasks were chosen based on their school events (e.g., athletic meetings, school festival) as well as on current social issues at the time of data collection (e.g., mayoral election), with a view to helping them develop their ability to talk about previous events in English. Three tasks which follow the same procedure with different contents were designed for the study (1) hosting an American friend; (2) describing school events; and (3) talking about mayor candidates. Whereas the task repetition group repeated the first task (i.e., hosting an American friend) three times, the procedural repetition group carried out all three tasks one time, each following the same procedure (i.e., describing previous events based on the pictures interactively, where one learner was in charge of the even numbered pictures, and the other learner was in charge of the odd numbered pictures).

Task complexity was manipulated using the [+/− reasoning demands] variable following Robinson's Cognition Hypothesis (2001). Taking the "Describing school events" task as an example, while the simple task group simply exchanged information about eight unique activities that they did at their schools (e.g., athletic meetings) with their partner, the complex task group was required to select one activity that should be continued and one activity that should not be continued in the following year while exchanging information. The process of evaluating and making a decision

about the information was hypothesized to increase the cognitive load. As shown in Table 1, task input that was provided to both groups was the same. All three tasks for procedural groups followed the same procedure (i.e., procedural repetition). In other words, the simple task group described (1) what they did with their American friends during their visit to Korea; (2) what students did during various school events; and (3) what each mayor candidate did over time based on given pictures. The complex group who was given the procedural repetition condition also repeated the same procedure three times, which involved making a decision while exchanging the information: (1) decide which event needs to be repeated or should not be repeated for next year's student exchange program; (2) decide which school event should be repeated or should not be repeated next year; and (3) decide which candidate should be elected or should not be elected for their city mayor.

Table 1. Descriptions of simple and complex version of "Describing school events" task

Simple	**Complex**
[**– reasoning demand**]	[**+ reasoning demand**]
Task input: A total of eight pictures which describe school events (e.g., athletic meetings) **Task outcome**: Report activities students did	**Task input**: A total of eight pictures which describe school events (e.g., athletic meetings) **Task outcome**: Suggest one event that they wish to continue next year and one event that they do not wish to continue.

Procedure

The tasks were administered during regular class hours for three days over a one-week period. During a 45-minute class period each day, learners carried out one task, which comprised planning time, the main task, and post-task activities. Because the beneficial effects of guided task planning have been documented in previous interaction studies (e.g., Foster & Skehan, 1996; Mochizuki & Ortega, 2008; Kim & McDonough, 2011), learners were shown two-minute-long task models during their planning time in the pre-task phase. During this phase, the researcher and the teacher demonstrated how to carry out the task to tacitly facilitate collaborative task performance and reinforce focus on form techniques. They used a different picture so that task modelling does not influence linguistic performance during tasks. The participants were then given 25–30 minutes to complete the main task. As mentioned in the previous section, while the exact task repetition group repeated the exact same task ("Hosting an American friend") three times, the procedural repetition group conducted all three information-exchange tasks, which differed in content but followed the same

procedure. Because it would be unlikely that the students repeat the exact same task three times without any change in real classroom contexts, and to help maintain attention and motivation, pairing of the participants in the task repetition group was changed for each task completion. By working on the same task with three different partners, learners repeated their output three times with the same materials; however, of course as a result of working with different partners each time, they received different input on the same topic on each occasion. In order to provide the same conditions with regard to potential effects of interlocutor change, the procedural repetition participants also changed their partners for each task.

Data analysis: Natural language processing tools (NLPT)

Students' spoken interactions were transcribed and then each transcript was segmented by speaker to produce two individual text files for each interaction. We used the Tool for the Automatic Analysis of Lexical Sophistication (TAALES; Kyle & Crossley, 2015) to calculate[1] three indices related to lexical sophistication: word familiarity, word age of acquisition, and word frequency. Such lexical sophistication indices have been widely used to account for the variance in lexical decision times (Kuperman, Stadthagen-Gonzales, & Brysbaert, 2012), lexical proficiency (e.g., Crossley et al., 2011a), speaking proficiency (e.g., Crossley & McNamara, 2013), and holistic scores of writing quality (e.g., Jung, Crossley, & McNamara, 2015; Kyle & Crossley, 2016). However, little to no research has examined whether task design or implementation variables affect L2 learners' lexical performance in terms of such indices.

Familiarity scores are based on judgments of how familiar words are to adults and are correlated with frequency counts (Crossley, Feng, Cai, & McNamara, 2013b). Words such as 'breakfast', 'radio', and 'book' all have high familiarity scores, while words such as 'encephalon' and 'egress' have low familiarity scores. TAALES includes familiarity scores for 4,943 lemmas reported by the MRC psycholinguistics database (Coltheart, 1981). Familiarity scores in the MRC database range from 43 (for 'mohul') to 657 (for 'breakfast'), with a mean of 488 (e.g., 'asphalt', 'bench', and 'destruction'). We selected the familiarity index that is calculated for all words in students' oral performance transcripts (including both function and content words).

Age of acquisition (AoA) scores are based on human judgments of the age that a particular L1 word is learned. We selected an AoA index from TAALES for all words based on the Kuperman et al. (2012) AoA list, which includes 30,121 lemmas that were judged for AoA using crowd sourcing techniques (see Kuperman et al. 2012, for

1. TAALES calculates average text scores based on a number of frequency and behavioral databases. For any given index, only the words in a target text that also occur in the database are counted towards the average score.

additional details). AoA scores collected by Kuperman et al. range from 1.58 years (for 'momma') to 25 years (for 'eisteddfod') with a mean of approximately 10 years (e.g., 'computing', 'degree', and 'aspect').

Frequency as a measure of lexical sophistication has been widely used in previous research. It has been shown to affect lexical decision times (Kuperman et al., 2012), indicating that high frequency words are processed more quickly than low frequency words. Word frequency indices have been found to be predictive of holistic quality/proficiency scores in L1 and L2 writing research (e.g., Crossley, Cobb, & McNamara, 2013a; Laufer & Nation, 1995). TAALES calculates frequency scores using the SUBTLEXus corpus (Brysbaert & New, 2009), which is comprised of subtitles from 8,388 films and television series from the United States (for more detailed information see Kyle & Crossley, 2015). Each word in this corpus is given a frequency norm for the number of times it occurs. For this study, we used logarithm transformed frequency scores for all words in the corpus, which control for the disproportional frequency of very frequent words (i.e., Zipfian distributions, Zipf, 1935).[2] We chose these three measures because they are theoretically linked to word exposure and word knowledge and, at the same time, previous studies have shown that the three indices are not strongly collinear (Crossley et al., 2011a, 2011b). Furthermore, these lexical sophistication measures are different from commonly used lexical diversity measures such as type/token ratio and lexical diversity (D) in that they show the properties of the actual words produced by learners independently from text internal factors such as text length.

Statistical analysis

A 3x4 mixed factorial analysis of variance (ANOVA) was conducted on the data. The between-groups variable was the task repetition conditions mediated by task complexity (simple task/procedure repetition, simple task/exact task repetition, complex task/exact task repetition, complex task/procedure repetition) and the within-groups variable was the three data collection points (i.e., time). We used the mixed factorial ANOVA to examine differences as a function of task repetition type (i.e., exact task repetition and procedural repetition) mediated by task complexity, differences as a function of time (i.e., the comparison between learners' performances across the three data collection points), and interactions between task and time. Prior to analyses, the data were checked for assumptions of normality and sphericity, which were met.

2. Because the frequency indices have been transformed, it is often difficult to interpret the reported numbers because they are not raw frequencies.

Results

Word familiarity

Descriptive statistics for the word familiarity analyses are reported in Table 2. Words with lower familiarity scores are considered more sophisticated. There was a significant and linear main effect for time on word familiarity, $F(2,69) = 23.038$, $p < .001$, $\eta_p^2 = .250$, such that the average measure of word familiarity decreased as the participants repeated more tasks (three times in total). In particular, Bonferroni post-hoc analyses revealed that the learners produced significantly greater numbers of less-familiar words at time 3 than at time 1 and time 2. The results also suggested a significant main effect for task implementation conditions, $F(3,69) = 8.610$, $p < .001$, $\eta_p^2 = .258$. Bonferroni post-hoc analyses showed that the learners who repeated exactly the same simple task three times produced greater numbers of familiar words than the rest of the groups. There was also a significant interaction between time and task implementation conditions, $F(6,69) = 5.073$, $p < .001$, $\eta_p^2 = .181$, indicating that word familiarity scores showed linear trends with time in the simple and complex tasks with procedural repetition, but not in the exact task repetition condition (see Figure 1). Between the simple and complex tasks, overall, complex tasks elicited less familiar words.

Table 2. Descriptive statistics for word familiarity: Mean (SD)

Group	Time 1	Time 2	Time 3
Simple/exact task repetition ($n = 11$)	599.231 (2.343)	599.768 (2.782)	598.83 (1.968)
Simple/procedure repetition ($n = 16$)	596.670 (4.734)	594.288 (4.689)	592.734 (2.793)
Complex/exact task repetition ($n = 24$)	595.833 (3.096)	594.623 (4.955)	595.499 (4.157)
Complex/procedure repetition ($n = 22$)	597.604 (2.443)	597.565 (1.719)	592.098 (4.384)
All tasks	597.062 (3.403)	596.211 (4.303)	594.37 (4.35)

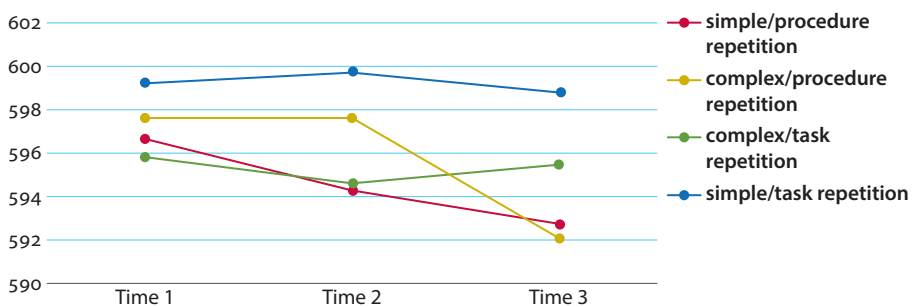

Figure 1. Interactions of word familiarity by task repetition and task complexity

Word age of acquisition

Descriptive statistics for the word age of acquisition analyses are reported in Table 3. Age of acquisition values represent the average age (in years) at which a word is learned by L1 speakers. As values increase, words are considered more sophisticated. There was a significant and linear main effect for time on age of acquisition, $F(2,69) = 94.050$, $p < .001$, $\eta_p^2 = .577$, such that word age of acquisition increased as the participants repeated tasks. In particular, Bonferroni post-hoc analyses revealed that the average word age of acquisition was significantly higher at time 3 than at time 2 and 1 as with the familiarity score. There was also a significant main effect for task implementation conditions, $F(3,69) = 15.434$, $p < .001$, $\eta_p^2 = .402$, such that both procedural repetition groups led to the production of words with higher age of acquisition than task repetition groups regardless of task complexity levels. There was also a significant interaction between time and task implementation conditions, $F(6,69) = 13.246$, $p < .001$, $\eta_p^2 = .365$, indicating that age of acquisition scores showed greater linear trends with time in the simple and complex procedural repetition groups but not in the simple and complex exact task repetition groups (see Figure 2).

Table 3. Descriptive statistics for word age of acquisition: Mean (SD)

Group	Time 1	Time 2	Time 3
Simple/exact task repetition ($n = 11$)	4.566 (0.207)	4.558 (0.153)	4.632 (0.124)
Simple/procedure repetition ($n = 16$)	4.729 (0.195)	4.807 (0.296)	5.231 (0.199)
Complex/exact task repetition ($n = 24$)	4.643 (0.167)	4.654 (0.263)	4.667 (0.234)
Complex/procedure repetition ($n = 22$)	4.681 (0.148)	4.729 (0.19)	5.152 (0.24)
All task conditions	4.661 (0.178)	4.696 (0.246)	4.931 (0.342)

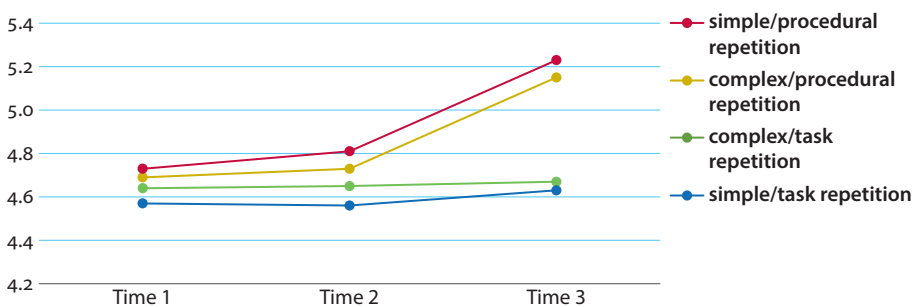

Figure 2. Interactions of word age of acquisition by task repetition and task complexity

Word frequency

Descriptive statistics for the word frequency analyses are reported in Table 4. Words that are less frequent are more difficult to process and/or learn and are considered more sophisticated. There was a significant and linear main effect for time on word frequency, $F(2,69) = 16.441$, $p < .001$, $\eta_p^2 = .192$, such that word frequency decreased as the participants repeated more tasks, as was also found for word familiarity and AoA. There was also a significant main effect for the task implementation factor, $F(3,69) = 8.182$, $p < .001$, $\eta_p^2 = .262$. Bonferroni post-hoc analyses revealed that the two procedural repetition groups led to the production of greater numbers of infrequent words as compared to the task repetition groups, particularly with simple tasks (see Figure 3). Unlike the results for word familiarity and AoA, a significant interaction between time and task implement condition was not reported for word frequency scores, $F(6,69) = 1.854$, $p > .050$, $\eta_p^2 = .075$.

Table 4. Descriptive statistics for word frequency: Mean (SD)

Group	Time 1	Time 2	Time 3
Simple/exact task repetition ($n = 11$)	4.805 (0.133)	4.789 (0.127)	4.700 (0.062)
Simple/procedure repetition ($n = 16$)	4.573 (0.139)	4.563 (0.187)	4.517 (0.099)
Complex/exact task repetition ($n = 24$)	4.710 (0.126)	4.677 (0.21)	4.708 (0.128)
Complex/procedure repetition ($n = 22$)	4.697 (0.147)	4.679 (0.177)	4.570 (0.169)
All tasks	4.69 (0.152)	4.67 (0.194)	4.623 (0.151)

Note. A log-transformed frequency score of 4.517 translates to a frequency of 636 occurrences per million words, while 4.805 translates to a frequency of 1243 instances per millions words

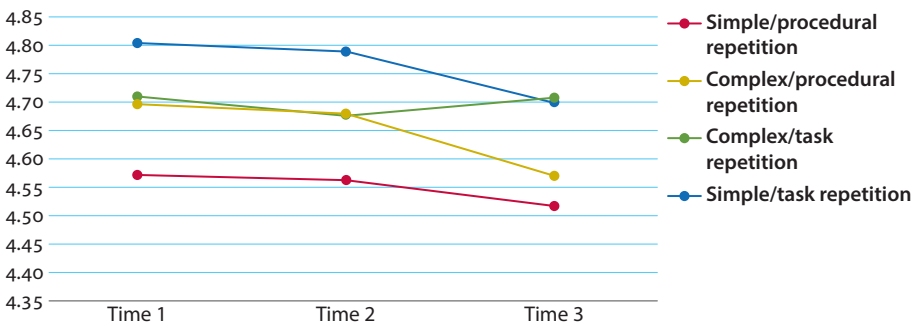

Figure 3. Interactions of word frequency by task repetition and task complexity

Discussion

The purpose of the current study was to investigate the impact of different types of task repetition on students' use of L2 vocabulary, and how such relations are mediated by task complexity. In terms of task implementation conditions, a total of four conditions were compared: simple/exact task repetition, simple/procedural repetition, complex/exact task repetition, and complex/procedural repetition. The three constructs focused on in the current study (i.e., word familiarity, word age of acquisition, and word frequency) all relate to the complexity of vocabulary and are variables that have not been used in previous task repetition and complexity studies. These variables differ from commonly employed lexical diversity indices such as type-token ratios in that they measure lexical properties of words (as compared to just ratios of types to tokens), they are not correlated with text length, and they are not associated with measures of text cohesion such as lexical overlap (McCarthy & Jarvis, 2010). In other words, lexical sophistication measures focus on text-external features of the lexical items in task performance data (e.g., oral interaction data in the current study). Furthermore, these indices provide information about the relative difficulty of learning the words in language production during task performance. Therefore, the variables provide us more potentially useful information about how task repetition and complexity can influence learners' ability to access and use more or less complex lexical features when compared to lexical diversity values. The fact that the indices included in the current study are strongly related to language proficiency and acquisition (Crossley et al., 2011b; Salsbury, Crossley, & McNamara, 2011) is valuable additional information that TBLT researchers can use when examining the role of task design and implementation features on the potential of task-based learning.

Overall, the results show that there were significant time effects for all three measures. In other words, as students repeated the exact same task or the same task procedure, students tended to produce words that were more sophisticated. Specifically, as the learners repeated either the exact same task or the same procedure, they were able to produce less frequent, less familiar words with higher age-of-acquisition scores. Researchers have suggested that greater lexical proficiency is characterized by the production of words that are not easily retrievable (i.e., less familiar; Crossley et al., 2011a) or are less frequent (Meara & Bell, 2001). Both more frequent and more familiar words are acquired earlier by learners in their language development because language teachers tend to put the most emphasis on such words (Nation, 2005). The current findings, then, may suggest that both exact task and procedural repetition help promote learners' production of less familiar, frequent words with higher age-of-acquisition scores, which might facilitate L2 lexical development in terms of more sophisticated word usage.

The benefits of repeating the exact same task in the production of more sophisticated vocabulary words as reported in this study are somewhat in line with Fukuta

(2016), whose findings suggest that repeating the exact same task could help learners free up their attentional resources away from focusing on the content, resulting in more diverse vocabulary use during the second enactment. Fukuta demonstrated that learners may be more likely to retain a memory of the conceptual and lexical aspects than formal aspects, and the semantic information may carry over to the following task more easily. More specifically, because learners become able to encode the words they encountered in the first task performance more rapidly via memory retrieval (Bygate & Samuda, 2005) as they repeat the same task twice, they are able to allocate spare attention and resources to the new vocabulary during the second enactment. Our finding seems to be consistent with Fukuta in this regard, indicating that repeating the exact same task may help learners develop their lexical proficiency (i.e., retrieving more sophisticated words from their lexicon), particularly in terms of their ability to use more sophisticated words (i.e., less frequent and familiar words with higher age-of-acquisition scores). However, with regard to procedural repetition, whereas Fukuta did not show any differences between the two task performances (same procedure with different content) on lexical diversity, the current study showed a significant change in lexical sophistication, as learners repeated the same task procedure with different content.

The differences found in the current study could also be interpreted in terms of monologic versus dialogic conditions. Unlike previous studies which implemented monologic picture description or story retell tasks (Bygate, 2001; Fukuta, 2016), the current study used collaborative tasks, which involved a more complicated procedure of task performance and discourse. This is particularly true for complex tasks which required learners to evaluate events/mayoral candidates and to make a decision after discussion. Since learners performed collaborative tasks in pairs, they were able to ask each other questions regarding unknown vocabulary items (i.e., lexical LREs). Previous research has shown that learners collaborate with each other in the solution of their linguistic problems and the co-construction of new language knowledge and such co-constructed knowledge tends to be retained by learners (Ellis, 2001; Lapkin, Swain, & Smith, 2002). More specifically, learners collaborate to solve lexical problems by pooling their linguistic resources and engaging in language-mediated cognitive activities such as offering and assessing new input, or formulating and testing hypotheses about the meaning and function of the words they are using (Dobao, 2012). It may be plausible that such negotiation may have promoted the use of unfamiliar and more sophisticated vocabulary words in the subsequent repeated task performance and led to building new lexical knowledge.

Furthermore, as Skehan (2009a) indicated, interaction occurring in a dialogic task is likely to confer benefits on the retrieval of lexical items during oral production. Such benefits include more time allocated to the process of lexical retrieval and potentially useful scaffolding and priming opportunities that may take place between

interlocutors as they work on the task collaboratively (Skehan, 2009b). Additionally, during interaction, learners' choice of lexical items could have been influenced by their interlocutors (i.e., through interactive alignment wherein speakers' utterances are influenced by recent discourse, Pickering & Garrod, 2004). Therefore, the findings of this study, which used dialogic tasks, cannot be directly compared to studies using monologic tasks without using caution in the interpretation of the results.

In terms of the mediating role of task complexity in the current study, complex tasks elicited less familiar words, words with higher age of acquisition, and less frequent words than simple tasks, particularly in the exact task repetition conditions (simple/exact task repetition vs. complex/exact task repetition). This finding is in line with some previous task complexity studies which have suggested that carrying out more complex tasks can facilitate lexical complexity during monologic oral task performance (e.g., Michel, Kuiken, & Vedder, 2007; Robinson, 2007; see Jackson & Suethanapornkul, 2013 for a meta-analysis). However, this pattern was not observed in the procedural repetition condition.

An interesting finding to note is that simple procedural repetition was strikingly more helpful in promoting the use of infrequent words than complex procedural repetition. This finding could be accounted for by the interaction effects between task complexity and different types of task repetition. Although while repeating the exact same complex version students produced more lexically sophisticated words compared to the simple version, when they had to repeat tasks with the same procedure but different content, complex versions of tasks might overload the processing burden. This in turn suggests that the procedural repetition condition with the simple task was more beneficial than with the complex task in terms of facilitating more sophisticated lexical words.

The results of the study may be interpreted in terms of different levels of cognitive demands associated with word frequency. Scholars have suggested that more frequent words tend to allow for faster decoding (Perfetti, 1985; Rayner & Pollatsek, 1994) and that frequent words are processed more quickly and comprehended better than infrequent words (Haberlandt & Graesser, 1985; Just & Carpenter, 1980). In other words, the processing burden may be alleviated when producing more frequent words, whereas greater cognitive demands are imposed for less frequent words. Considering an L2 learner's mental lexicon is not fully organized or elaborated, we assume that learners were able to fulfill the need to access more difficult lemmas while repeating exactly the same task, particularly favoring more complex version of the task. On the other hand, repeating more complex tasks with greater cognitive demands caused by new content may have led learners to produce more frequent (i.e., less difficult) lemmas (Kormos, 2006; Skehan, 2009b). To date, many task complexity studies have suggested that more complex tasks promote a greater amount of lexical LREs (e.g., Kim, 2013); however, these studies did not shed light on the details about

lexical sophistication of target words being discussed during LREs. The current study suggests that when multiple tasks are performed to examine the role of task complexity, task repetition condition also needs to be carefully controlled.

The current study has several implications for research methods and language pedagogy in task repetition research. First of all, we find that task repetition leads to the production of more sophisticated words and that repeating simple tasks, as compared to complex tasks, also leads to the production of more sophisticated vocabulary. Secondly, the current study used automated measures of lexical sophistication to assess how task repetition and complexity conditions can influence the production of L2 vocabulary during collaborative task performance. To our knowledge, it is the first study to examine lexical sophistication in interaction data using natural language processing tools, with most studies to date having used hand coding (Plonsky & Kim, 2016). In order to advance our understanding of task-based learner production, future studies would probably benefit from using natural language processing or corpus tools.

The current study also sheds insights on the value of using task repetition in language classrooms. For one thing, the results suggest the benefits of task repetition particularly for enabling learners to use less frequent words and less familiar words. Even repeating the exact same task could facilitate the production of more sophisticated words. In terms of the role of task complexity as a mediating variable, the current study did not show significant task complexity effects. Of interest is the notion that the effects of task repetition may be impacted by task complexity as observed in the interaction effects reported in the study. For example, the current findings demonstrated that word familiarity and age of acquisition scores showed greater linear trends with time (i.e., the three data collection points) in the simple and complex procedural repetition groups but not in the simple and complex task repetition groups.

Future research should also address the limitations of the current study. In order to examine the role of repeating the procedure of tasks only, it was necessary to use the same task types with different content which introduced the potential for topic-based effects. Previous research has reported that task topic can exert a significant influence on L2 learner's task performance in terms of lexical complexity in that different task topics may call upon different levels of lexical complexity (Skehan, 2014; Vermeer, 2000; Yu, 2009). Since the study was conducted during the participants regular English classes, we were not able to counter-balance the order of topics in the procedural repetition conditions, which could have addressed the methodological concern regarding the intervening role of topic. Future studies may want to counterbalance the order of topics in order to control the topic effects when examining the procedural repetition condition. Additionally, the current study examined students' task performance only. Future research is needed to examine the role of different aspects of task repetition on the long-term learning of L2 vocabulary words.

Conclusion

Building on previous task repetition research, the current study examined the effects of repeating different aspects of tasks (i.e., exact task repetition vs. procedural repetition) on the lexical production of paired language learners as they repeated collaborative tasks. Learners' use of L2 vocabulary was measured in terms of word familiarity, word age of acquisition, and word frequency, and the impact of different levels of task complexity (i.e., complex vs. simple task) in mediating task repetition effects was also investigated, however showing null results.

The findings of the current study address both methodological and pedagogical implications. L2 instructors might want to implement task repetition in class but consider different complexity levels for different task repetition conditions. In general, procedural repetition might be an effective way to facilitate the use of more sophisticated words. Additionally, complex tasks might be more appropriate when the exact task repetition is applied. Methodologically, future task-based interaction studies could benefit from analyzing L2 learners' linguistic performance during collaborative tasks by using natural language processing tools. To date, task repetition research has offered useful guidelines for TBLT; however, many studies have been conducted in tightly controlled lab-settings. More classroom-based research is needed, and future research may also address different interventions along with task repetition (e.g., provision of corrective feedback).

References

Ahmadian, M. J. (2013). Working memory and task repetition in second language oral production. *Asian Journal of English Language Teaching*, 23, 37–55.

Ahmadian, M.J., & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research*, 15, 35–59. https://doi.org/10.1177/1362168810383329

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41, 977–990. https://doi.org/10.3758/BRM.41.4.977

Bygate, M. (1996). Effects of task repetition: Appraising the developing language of learners. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 136–146). London: Heinemann.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 23–48). Harlow: Longman.

Bygate, M. (2016). Sources, developments and directions of task-based language teaching. *The Language Learning Journal*, 44, 381–400.

Bygate, M., & Samuda, V. (2005). Integrative planning through the use of task repetition. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 37–74). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.11.05byg

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology*, 33, 497–505. https://doi.org/10.1080/14640748108400805

Crossley, S. A., Cobb, T., & McNamara, D. S. (2013a). Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications. *System*, 41, 965–981. https://doi.org/10.1016/j.system.2013.08.002

Crossley, S. A., Feng, S., Cai, Z., & McNamara, D. S. (2013b). Computer simulations of MRC Psycholinguistics Database word properties: Concreteness, familiarity, and imageability. In S. Jarvis & M. Daller (Eds.), *Vocabulary knowledge: Human ratings and automated measures* (pp. 135–156). Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.47.07ch5

Crossley, S., & McNamara, D. (2013). Applications of text analysis tools for spoken response grading. *Language Learning & Technology*, 17, 171–192.

Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011a). What is lexical proficiency? Some answers from computational models of speech data. *TESOL Quarterly*, 45, 182–193. https://doi.org/10.5054/tq.2010.244019

Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011b). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing*, 28, 561–580. https://doi.org/10.1177/0265532210378031

Dobao, A. F. (2012). Collaborative dialogue in learner–learner and learner–native speaker interaction. *Applied Linguistics*, 33, 229–256. https://doi.org/10.1093/applin/ams002

Ellis, R. (2001). Non-reciprocal tasks, comprehension and second language acquisition. In M. Bygate, P. Skehan, & M. Swain. (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 49–74). Harlow: Longman.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18, 299–323. https://doi.org/10.1017/S0272263100015047

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21, 354–375. https://doi.org/10.1093/applin/21.3.354

Fukuta, J. (2016). Effects of task repetition on learners' attention orientation in L2 oral production. *Language Teaching Research*, 20, 321–340.

Gass, S., Mackey, A., Alvarez-Torres, M. J., & Fernández-García, M. (1999). The effects of task repetition on linguistic output. *Language Learning*, 49, 549–581. https://doi.org/10.1111/0023-8333.00102

Guo, L. Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Writing Assessment*, 18, 218–238. https://doi.org/10.1016/j.asw.2013.05.002

Haberlandt, K. F., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology: General*, 114, 357–374. https://doi.org/10.1037/0096-3445.114.3.357

Hawkes, M. (2012). Using task repetition to direct learner attention and focus on form. *ELT Journal*, 66, 327–336. https://doi.org/10.1093/elt/ccr059

Jackson, D. O., & Suethanapornkul, S. (2013). The cognition hypothesis: A synthesis and meta-analysis of research on second language task complexity. *Language Learning*, 63, 330–367.

Jung, Y., Crossley, S. A., & McNamara, D. S. (2015). Linguistic features in MELAB writing task performance (*Working Paper No. 2015–05*). Retrieved from Cambridge Michigan Language Assessment website: <http://www.cambridgemichigan.org/about-us/research/working-papers/>

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329–354. https://doi.org/10.1037/0033-295X.87.4.329

Kim, Y. (2013). Promoting attention to form through task repetition in. In K. McDonough & A. Mackey (Eds.), *Second language interaction in diverse educational contexts* (pp. 3–24). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.34.04ch1

Kim, Y. (2015). The role of tasks as vehicles for learning in classroom interaction. In N. Markee (Ed.), *Handbook of classroom discourse and interaction* (pp. 163–181). Malden, MA: Wiley-Blackwell.

Kim, Y., & McDonough, K. (2011). Using pre-task modeling to encourage collaborative learning opportunities. *Language Teaching Research*, 15, 1–17. https://doi.org/10.1177/1362168810388711

Kim, Y., & Payant, C. (2014). A pedagogical proposal for task sequencing: An exploration of task repetition and task complexity on learning opportunities. In M. Baralt, R. Gilabert, & P. Robinson (Eds.), *Task sequencing and instructed second language learning* (pp. 151–177). London: Bloomsbury Academic.

Kim, Y., & Tracy-Ventura, N. (2013). The role of task repetition in L2 performance development: What needs to be repeated during task-based interaction? *System*, 41, 829–840. https://doi.org/10.1016/j.system.2013.08.005

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kuperman, V., Stadthagen-Gonzales, H, & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44, 978–990. https://doi.org/10.3758/s13428-012-0210-4

Kyle, K. & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, 49, 757–786. https://doi.org/10.1002/tesq.194

Kyle, K. & Crossley, S. A. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing*, 34, 12–24.

Lapkin, S., Swain, M., & Smith, M. (2002). Reformulation and the learning of French pronominal verbs in a Canadian French immersion context. *The Modern Language Journal*, 86, 485–507. https://doi.org/10.1111/1540-4781.00157

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16, 307–322. https://doi.org/10.1093/applin/16.3.307

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

Long, M. (2015). *Second language acquisition and task-based language teaching*. Malden, MA: Wiley Blackwell.

Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, 4, 221–250.

Lynch, T. (2007). Learning from the transcripts of an oral communication task. *ELT journal*, 61, 311–320. https://doi.org/10.1093/elt/ccm050

Mackey, A. (2007). Interaction as practice. In R. M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 85–110). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511667275.006

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior research methods*, 42, 381–392.

Meara, P., & Bell, H. (2001). P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect*, 3, 5–19.

Michel, M. C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *International Review of Applied Linguistics in Language Teaching*, 45, 241–259. https://doi.org/10.1515/iral.2007.011

Mochizuki, N., & Ortega, L. (2008). Balancing communication and grammar in beginning-level foreign language classrooms: A study of guided planning and relativization. *Language Teaching Research*, 12, 11–37. https://doi.org/10.1177/1362168807084492

Nation, P. (2005). Second language vocabulary. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (Vol. 14, pp. 413–448). Amsterdam: Elsevier.

Patanasorn, C. (2010). *Effects of procedural, content, and task repetition on accuracy and fluency in an EFL context* (Unpublished doctoral dissertation). Northern Arizona University, Flagstaff, AZ.

Perfetti, C. A. (1985). *Reading ability*. Oxford: Oxford University Press.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, 169–190. https://doi.org/10.1017/S0140525X04000056

Plonsky, L., & Kim, Y. (2016). Task-based learner production: A substantive and methodological review. *Annual Review of Applied Linguistics*, 36, 73–97.

Rayner, K, & Pollatsek, A. (1994). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.

Robinson, P. (2001). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287–318). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139524780.012

Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *International Review of Applied Linguistics in Language Teaching*, 45, 193–213. https://doi.org/10.1515/iral.2007.009

Salsbury, T., Crossley, S. A., & McNamara, D. S. (2011). Psycholinguistic word information in second language oral discourse. *Second Language Research*, 27, 343–360. https://doi.org/10.1177/0267658310395851

Skehan, P. (2009a). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30, 510–532. https://doi.org/10.1093/applin/amp047

Skehan, P. (2009b). Models of speaking and the assessment of second language proficiency. In A. Benati (Ed.), *Issues in second language proficiency* (pp. 203–215). London: Continuum.

Skehan, P. (2014). Limited attentional capacity, second language performance, and task-based pedagogy. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 211–260). Amsterdam: John Benjamins. https://doi.org/10.1075/tblt.5.08ske

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *The cModern Language Journal*, 82, 320–337.

Van den Branden, K., Bygate, M., & Norris, J. M. (2009). Task-based language teaching: Introducing the reader. In K. Van den Branden, M. Bygate & J. M. Norris (Eds.), *Task-based language teaching: A reader* (pp. 1–13). Amsterdam: John Benjamins.

Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing*, 17, 65–83. https://doi.org/10.1177/026553220001700103

Yu, G. (2009). Lexical diversity in writing and speaking task performances. *Applied Linguistics*, 31, 236–259. https://doi.org/10.1093/applin/amp024

Zipf, G. K. (1935). *The psycho-biology of language*. New York, NY: Houghton-Mifflin.

CHAPTER 4

Discourse performance in L2 task repetition

Zhan Wang and Gaowei Chen
South University of Science and Technology of China / University of Hong Kong

'Repetition' in language practice has often been associated with behavioristic drills, typically regarded as giving rise to habit formation, and often excluded from mainstream language teaching approaches. However, this study, drawing upon Bygate's (1999) argument that task is a context for the framing and reframing of language, shows that repetition in meaningful tasks does not necessarily involve passive performance repetition; instead, L2 learners are actively re-constructing the narrative discourse. This study provides support to this argument through the analysis of the speech samples from thirteen learners performing an immediate exact L2 task repetition. The participants were asked to perform a video-based monologue story narration twice to an imagined listener, and the comparison of the learners' first and second speech performances shows that the learners enhanced their discourse performance with: (1) increased quantity of discourse (evidenced by more Total Words, Total Propositions, and higher P-density); (2) increased cohesion (evidenced by higher ratio of given words to new words in sentences based on latent semantic analysis, i.e., LSA Given/New value); and (3) increased quality of lexis (evidenced by higher Noun Hyperonymy value, indicating more specific nouns used). The results suggest that task repetition, in addition to its effects on language complexity, accuracy, and fluency, has an impact on learners' speech performance at the discourse level.

Introduction

Research on task-based language teaching and learning has developed significantly in recent years (Ellis, 2003, 2005; Long 2014; Robinson, 2011; Skehan, 2009; Van den Branden, Bygate, & Norris, 2009). Research findings in general lend support to the effectiveness of using 'tasks', in which meaning is primary, with variations of conditions and characteristics for the improvement of second language performance and proficiency (Samuda & Bygate, 2008; Skehan, Bui, Li, & Wang, 2012). However, among the task conditions, task repetition is perhaps the least studied (though notable studies include Ahmadian & Tavakoli, 2011; Bygate, 1996, 1999, 2001; De Jong & Perfetti, 2011; Fukuta, 2016; Gass, Mackey, Alvarez-Torres, & Fernandez-Garcia, 1999; Lynch

DOI 10.1075/tblt.11.04wan

& Maclean, 2000; Qiu & Lo, 2017). On one hand, we have far less knowledge about the basic types of repetition (e.g., repetition times and intervals) and their effects on language performance. On the other hand, language performance in tasks, commonly measured with various indices of Complexity, Accuracy, Lexis and Fluency (CALF) (Ellis, 2009; Housen & Kuiken, 2009; Norris & Ortega, 2009; Skehan, 2009), is still in need of research into other indices especially at the discourse level. If we can find more evidence about whether and how task repetition could impact on language performance at the discourse level, this would shed more light on the value of task repetition for language learning and teaching, and suggest implications for the effective design and use of tasks.

Effects of task repetition on speech performance (CALF)

Task repetition as an option within TBLT refers to the learner repeating a task at different times (Ellis, 2003; Samuda & Bygate, 2008; Skehan, 1998). Bygate (1996, 1999, 2001) argues that for holistic meaning-focused tasks, repetition will not lead to verbatim repetitive performance but to learners re-constructing their interlanguage each time, using the opportunity to frame and reframe their language. Evidence for the claim of effects on speech performance can be found in the literature in terms of three variations in task repetition: (1) immediate task repetition, (2) task repetition after a time interval, and (3) multiple task repetitions with time intervals (see Bui, Skehan, & Wang, in press).

Findings from these domains of research indicate that, first, the effects of *immediate task repetition* on the enhancement of speech complexity, accuracy, and fluency (CAF), though not on lexical performance, are generally consistent. For example, in a study of immediate monologue task repetition (i.e., repeating the task to an imagined listener), Wang (2014) found enhanced speech complexity, accuracy, and fluency in the learners' second speech performance. In studies of multiple immediate repetitions, Lynch and Maclean (2000) found improved accuracy and fluency in the speech performance of L2 leaners, who presented a poster six times to different interlocutors. Lambert, Kormos, and Minn (2017) found that multiple immediate task repetitions effectively enhanced speech fluency on various measures, although different measures began showing changes in different rounds of repetition.

Second, research findings on the effects of *task repetition after a time interval*, usually involving an interval of two-three days to one week, are also consistent. This type of task repetition has shown effects mainly on the enhancement of speech accuracy and fluency (AF) but less so on lexis. In a single case study, Bygate (1996) found that repeating the same task after a three-day interval helped to improve speech accuracy and fluency for an intermediate-proficiency learner. Similarly, Ahmadian and

Tavakoli (2011) found that, after a one-week interval, intermediate learners improved speech complexity, accuracy, and fluency; and Fukuta (2016) found improved speech accuracy and lexical performance (measured by Guiraud's index) with upper-intermediate learners.

A third type of task repetition is *multiple task repetitions with time intervals*, which involves repeating the same specific task or the same type of task multiple times usually over periods longer than one week. In these studies, research findings are mixed. For example, Bygate (2001) investigated the effects of repetition of an exact same task and multiple repetitions of the same task type on L2 speech performance, and found that the effects of an initial enactment of a task on the speech complexity and fluency (but not on accuracy) of learners could be found on performance of the same task 10 weeks later. He found however no effect for task type repetition. In an investigation of task repetition involving three different training conditions across two weeks: an exact same task repetition, a task type repetition, and a control group without repetition training, Gass et al. (1999) found that the exact task repetition group outperformed the control group on general proficiency, on accuracy of a Spanish morphosyntactic element, on lexical density, and on lexical sophistication (i.e., the number of difficult words used). And repeating the same task but with increasing time pressure on each iteration, de Jong and Perfetti (2011, see also this volume) found that, as a training method, the 4–3–2 task repetition (i.e., repeating a topic for four, three, and two minutes with increased time pressure) could significantly enhance the speech fluency of participants compared to a control group which spoke on three different topics each time.

However, studies such as these limited their focus to the impact of task repetition on CALF measures, with little interest so far in the impact of task repetition on discourse performance (though see Bygate & Samuda, 2005). Yet TBLT is an approach with the potential to improving discourse competence, and this is the focus of the present chapter.

Discourse competence in language teaching, learning, and assessment

Discourse competence

Discourse competence is the ability to arrange sentences or utterances in sequence so as to produce coherent language (Council of Europe, 2001, p. 123). It is a concept often included in communicative language teaching models. For example, Canale and Swain (1980) and Canale (1983) distinguish four aspects of communicative competence: (1) grammatical/linguistic competence (the knowledge of lexis, syntax, semantics, and phonology); (2) sociolinguistic competence (the sociocultural rules under

communication); (3) strategic competence (the use of strategies, both verbal and nonverbal, when communication breaks down); and (4) discourse competence (the cohesion and coherence of sentences and utterances). The concept of discourse competence is also included in Bachman (1990) and Bachman and Palmer's (1996) models of language ability, which consists of organizational competence (i.e., grammatical and textual competence), pragmatic competence (i.e., functional and sociolinguistic competence), and strategic competence. In their model, discourse knowledge and ability is incorporated into organizational competence and pragmatic competence.

Discourse competence in language teaching and learning

Discourse has been a focus in language teaching and learning since the 1970s. Widdowson (1978) for example proposed, "Since our aim is to get the learner to cope with discourse in one way or another, it would seem reasonable to suggest that instances of discourse should serve as the point of reference for all the exercises which are devised" (1978, p. 146). He suggested that teaching units and teaching tasks should be instances of discourse scaffolding, linking one to another (e.g., from receptive tasks such as reading passages to productive tasks such as writing passages). Other discourse-relevant approaches to language teaching and learning can be found in Cook (1989), Nunan (1989), and McCarthy & Carter (2014), suggesting a variety of activities to help implement discourse-based pedagogy.

Discourse performance in language assessment (CEFR)

The Common European Framework of Reference (CEFR) for Languages also includes discourse competence in its approach. It proposes three-component model of communicative language competence (i.e., linguistic, sociolinguistic, and pragmatic competences) to measure L2 proficiency, with discourse competence seen as part of the umbrella component of pragmatic competence (Council of Europe, 2001). Operationally, it identifies the criterion of coherence as one of the five aspects (i.e., range, accuracy, fluency, interaction, and coherence) used to assess spoken language performance and judge the learner's proficiency level. The rubrics describing the common reference levels under the coherence criterion are listed in Table 1 (cited from Council of Europe, 2001, p. 37).

Table 1 shows how the CEFR uses different degrees of 'coherence' and 'cohesion' in the description of the rubrics to evaluate discourse competence. While these criteria are intended for human raters to judge the speakers' language performance, are there automatic or more objective methods for assessing discourse performance? The CEFR suggests some other indices to measure the ability of language use at discourse level (e.g., topic/focus, given/new, cause/effect, thematic organization, coherence and

Table 1. Common reference levels for the coherence aspect of spoken language use in the CEFR (Council of Europe, 2001, p. 37)

	Coherence
C2	Can create coherent and cohesive discourse, making full and appropriate use of a variety of organisational patterns and a wide range of connectors and other cohesive devices.
C1	Can produce clear, smoothly flowing, well-structured speech, showing controlled use of organisational patterns, connectors and cohesive devices.
B2	Can use a limited number of cohesive devices to link his/her utterances into clear, coherent discourse, though there may be some 'jumpiness' in a long contribution.
B1	Can link a series of shorter, discrete simple elements into a connected, linear sequence of points.
A2	Can link groups of words with simple connectors like 'and', 'but' and 'because'.
A1	Can link words or groups of words with very basic linear connectors 'and' or 'then'.

cohesion, Council of Europe, 2001, p. 123). These discourse indices, many of which have been integrated into software programs by computational linguists in recent years, could expand upon the existing CALF measures to include new discourse aspects. What is clear is that there is a broad consensus that discourse competence needs to be included in studying language development, which of course implies that it has a valuable contribution to make in TBLT research.

Discourse performance measures in this study

In this study, following the CEFR framework, we tentatively use indices automatically produced by computer to measure three sub-dimensions of discourse performance: (1) discourse quantity, (2) cohesion, and (3) lexical quality. There are two reasons for our selection of these discourse dimensions. First, for a pragmatic reason, free computer programs are available for producing the indices of these measures automatically (e.g., CPIDR and Coh-Metrix, to be introduced in *Methodology*). Second, but more importantly, we use them as meaningful measures of L2 narrative discourse because they are most likely direct products of the conceptualization and lexical selection processes in speech production (Levelt, 1989). The conceptualization stage involves macro- and micro-planning, with the former concerned with a discourse focus, involving attention on whether what should be said will be said, and whether the interlocutor is following the speech act (Levelt, 1999, p. 90). Micro-planning is concerned with the propositional content of the utterances to fulfil this purpose (Levelt, 1999, p. 93). We also assume that it is reasonable to expect these indices to be sensitive to the task repetition condition. Thus, the speakers should

benefit from the monitoring opportunity available to the speech comprehension system during their first speech performance, and thus gain more time and attentional space during their second speech performance. This would enable them to generate more sophisticated macro- and micro-planning based on their conceptualization needs, and conduct more precise lexical selection, mapping their conceptual plan to their mental lexicon.

These three dimensions of discourse performance are all relevant to the investigation of task repetition effects. As suggested by Bygate (1999), task repetition provides L2 learners with an opportunity to reconstruct their language by building on memory of their speech performance on previous occasions. And indeed Bygate and Samuda (2005) reported evidence of effects of task repetition on aspects of learners' narrative discourse. In comparison to other types of task repetition, immediate task repetition is probably less affected by "forgetting" due to the short interval of time between iterations. As noted above the literature shows consistent effects of immediate task repetition on learners' speech complexity, accuracy, and fluency (Lambert et al., 2017; Lynch & Maclean, 2000; Wang, 2014). Therefore, we use immediate task repetition to test if task repetition can enhance the L2 learners' narrative discourse performance in terms of the three above mentioned aspects: discourse quantity, cohesion, and lexical quality.

Research questions

This study investigated the following three research questions:

1. Does immediate task repetition result in a significant increase in the quantity of discourse?
2. Does immediate task repetition result in significantly improved cohesion?
3. Does immediate task repetition result in significantly improved lexical quality?

The study then aims at investigating the effects of immediate task repetition on learners' discourse. We measure discourse quantity by the total number of words, the total number of propositions, and the propositional idea density of speech samples. The measures for cohesion are the ratio of given words to new words in sentences based on latent semantic analysis (LSA, i.e., Given/New values), the incidence of causal connectives, and the ratio of causal particles to causal verbs. Lexical quality is measured by lexical diversity, word frequency, word familiarity, word concreteness, word imageability, and Hypernymic value. The nature of the specific measures is explained more fully in the methodology section.

The hypotheses underpinning these questions are two-fold. For the first and second research questions, we hypothesize that immediate task repetition will improve

discourse quantity and cohesion, but we are less confident that it will enhance lexical quality. The reason for this is that, although the later enactment of a narrative task provides the speaker with more time and attentional space to conduct macro- and micro-planning, additional time and attentional space may not compensate for the L2 learner's comparatively small lexicon in lexical selection. Especially with immediate task repetition, improving lexical quality of L2 speech is difficult, as shown by the inconsistent results of lexical performance in the task repetition literature. In contrast, as learners become more familiar with the content of the task, the amount and cohesion of the discourse are more likely to increase.

Methodology

Participants

The participants were 13 undergraduates recruited from a Hong Kong university as volunteers, all aged between 18 and 22, and majoring in different subjects. They were native Chinese speakers who learned English as a second language. At the time of the data collection, their self-reported English proficiency was based on the TOEFL Internet Based Test (iBT) and ranged from 76 to 109, and IELTS 6 to 7.5, with speaking scores ranging between 5.5 and 7.5. The scores can be interpreted as corresponding to the CEFR levels from B2 to C2, according to the Educational Testing Service (ETS) official website <https://www.ets.org/toefl/institutions/scores/compare/>. A pre-test showed that the participants' English proficiency was between intermediate and advanced levels (with the mean being 35.8, and a standard deviation (SD) of 5.3), using a mock research version of the TOEFL listening comprehension subtest from Hinkel (2004) (about 45 minutes for 50 multiple choice questions, and the full score being 50).

Immediate task repetition

The study used a video-based narrative task, in which a silent five-minute Mr. Bean video was played normally as a prompt to elicit, twice, the participants' narrative performances. A video entitled 'Shopping" showed Mr. Bean entering a shopping mall and behaving bizarrely before making a purchase (e.g., unpacking and trying a toothbrush, wrapping himself in a bath towel, and testing a frying pan for size with a fish). The video was from the same Mr. Bean video series used in Skehan and Foster (1999), and, together with the task procedure, was piloted before the study with participants of the same background and from the same university as those participating in the actual study. The pilot participants confirmed the clarity of the video content as well

as the feasibility of the immediate task repetition condition. While the video was playing normally, the participants told the story simultaneously as if they were providing a running commentary to the audience, which yielded their first speech performance. They were then told to narrate the story again, under the same condition, so that the learners narrated the story twice.

Data collection procedure

Data collection occurred through one-on-one meetings with participants. After the 13 participants had completed the pre-test – the mock TOEFL listening comprehension subtest – they were asked to watch the Mr. Bean video and simultaneously narrate the story to an imagined listener, which produced speech samples of their first performance. Immediately afterwards they were required to narrate the video for a second time, under the same conditions, which yielded speech samples of their second performance. Finally, the participants were interviewed by a researcher and completed a questionnaire about their English learning experience.

Measures of discourse performancest

Table 2 lists the specific indices for measuring the dimensions of discourse quantity, cohesion and lexical quality. These discourse performance dimensions were assessed primarily because of their connection with the conceptualization stage of speech

Table 2. Discourse performance measures in this study

Discourse quantity
– Total Words (from CPIDR)
– Total Propositions (from CPIDR)
– P-density: Propositional idea density (from CPIDR)
Cohesion
– LSA (Given/New, from Coh-metrix)
– Causal Connective Incidence (occurrence per 1,000 words) (from Coh-metrix)
– Causal Ratio (Ratio of causal particles to causal verbs) (from Coh-metrix)
Lexical quality
– Lexical Diversity (MTLD) (from Coh-metrix)
– Lexical Diversity (vocd-D) (from Coh-metrix)
– Word Frequency (CELEX log word frequency for all words, mean, via Coh-Metrix)
– Word Familiarity (for content words, ratings from MRC database via Coh-Metrix)
– Word Concreteness (for content words, ratings from MRC database via Coh-Metrix)
– Word Imageability (ratings from MRC database via Coh-Metrix)
– Hypernymic Value (for nouns, from WordNet database via Coh-Metrix)

production, as discussed earlier. Moreover, they can be produced automatically using existing software. The description for each index is provided in the next section.

Discourse quantity

Three indices were used (all drawn from the Computerized Propositional Idea Density Rater (CIPDR), Brown, Snodgrass, Kemper, & Herman, 2008) in measuring the dimension of discourse quantity: Total Words, Total Propositions, and Propositional idea density (P-density). *Total Words* is the total number of words produced in the learner's speech. This is often regarded as a measure of syntactic complexity or a fluency measure (especially for fixed-time speaking tasks). However, in the current study, total words (as well as total propositions and propositional idea density) are taken as quantitative indicators of discourse, resulting from macro planning during conceptualization. *Total propositions* refers to the total number of propositions, corresponding roughly to the verbs, adjectives, adverbs, prepositions, and conjunctions used (i.e., the content words but not the grammatical words). And the quotient of the Total Propositions and the Total Words is the *Propositional idea density*, which is taken as a measure of the density of content words in relation to the total number of words. Taken together, P-density, Propositions, and Total Words are used to reflect the amount of conceptualization and the effect of macroplanning (Levelt, 1999). For implementing these three measures of discourse quantity, we used the CPIDR (Brown, et al., 2008), which automatically provides the Total Words count, the Total Propositions count, and the P-density in the output.

Cohesion

Several cohesive devices are related to speech discourse coherence, such as those that indicate temporal reference, topic, givenness, newness, causal effect, thematic role, and organization, as suggested in the CEFR (Council of Europe, 2001, p. 123). For this study, we used three indices to measure cohesion: Latent Semantic Analysis (LSA Given/New), (Hempelmann et al., 2005; Landauer et al., 2007), Causal Connective Incidence, and Causal Ratio. *LSA Given/New* measures how much given versus new information there is in each sentence in a text using the formula "G/(N+G)", and more given information in a text is taken to indicate higher cohesion of the text (Hempelmann et al., 2005; McCarthy, Dufty, Hempelman, Cai, Graesser, & McNamara, 2012; McNamara, Graesser, McCarthy, & Cai, 2014). *Causal Connective Incidence* is the number of causal connectives (e.g., "because", "so"), and the *Causal Ratio* is the ratio of causal connectives to total words. These indices were used because we assume that they connect to both macro- and micro-planning during the speech conceptualization stage, and have been used in other studies to evaluate low-cohesion

versus high-cohesion texts in both spoken and written discourse (McCarthy et al., 2012; McNamara, Louwerse, McCarthy, & Graesser, 2010).

To count the three cohesive indices automatically, we used Coh-Metrix, a computational system developed at the University of Memphis (Graesser, McNamara, Louwerse, & Cai, 2004; Graesser, McNamara, & Kulikowich, 2011; McNamara, Louwerse, Cai, & Graesser, 2005, <www.cohmetrix.com>). Coh-Metrix also provides many other measures of cohesion, such as logical connectives (e.g., "and", "or"), adversative/contrastive (e.g., "although", "whereas"), temporal (e.g., "first", "until"), and additive (e.g., "and", "moreover") relations, but we did not use them for this study because, for the short narrative speech samples produced by the learners, we assumed that either there were too few incidences (e.g., adversative and contrastive) or too many pseudo connectives (e.g., 'and' used habitually in spoken discourse but not as real connectives; see Foster, Tonkyn, & Wigglesworth, 2000).

Lexical quality

While lexical quality ultimately reflects the language proficiency of L2 learners, it may reflect the real-time 'lexical selection' by the speaker as well; namely, selecting the appropriate words from the mental lexicon to map with the conceived information (Levelt & Indefrey, 2000, p. 79). Therefore, we assume that lexical quality should be used as one sub-dimension of discourse performance. For this study, we measured lexical quality using seven indices: MTLD (McCarthy, 2005), vocd-D (Malvern, Richards, Chipere, & Durán, 2004), Word Frequency, Word Familiarity, Word Concreteness, Word Imageability, and Hypernymic Value (for nouns) with the Coh-Metrix program (Graesser et al., 2004; McNamara et al., 2014). These indices are related to individual word properties and have been used in the literature to assess the lexical quality of L2 learners' oral and written discourse, as well as to compare the difference between words produced by L2 learners and by native speakers (Crossley & Salsbury, 2010; Crossley, Salsbury, & McNamara, 2010). To measure the discourse performance of L2 learners, the Coh-Metrix program was used to produce the indices of lexical quality. We now describe each of the seven indices in turn.

Lexical diversity (LD) is a lexical measure (often used interchangeably with 'lexical variation' and 'lexical variety' (see Engber, 1995; cited in Jarvis, 2013). Usually, LD assesses the proportion of different words used in a text (McCarthy & Jarvis, 2010; Jarvis, 2013), and an often-heard index of lexical diversity is the 'type-toke ratio', which divides the number of word 'types' by the total word count or 'tokens' (TTR, Templin, 1957). However, researchers have pointed out that, because it is sensitive to text length, TTR cannot be a reliable measure, that is, the longer the text, the lower the ratio (Malvern et al., 2004). Adjustments have been made in other lexical diversity indices, such as MTLD (McCarthy, 2005), vocd-D (Malvern et al., 2004), and

HD-D (hypergeometric distribution). McCarthy and Jarvis (2010), who validated these different approaches to assessing lexical diversity, found that MTLD performs well with respect to various validity criteria, and did not vary as a function of text length. Meanwhile they suggested that researchers use these indices rather than any single one of them, as they capture unique lexical information. Following their advice, we used MTLD and vocd-D together to assess lexical diversity in the speech samples.

Word Frequency value can be understood, albeit roughly, as the figure that represents the frequency for a given word-form or a lemma per million in language databases such as CELEX (Baayen, Piepenbrock, & Gulikers, 1995, <celex.mpi.nl>) or the British National Corpus 2014 (Love, Dembry, Hardie, Brezina, & McEnery, 2017, <http://corpora.lancs.ac.uk/bnc2014>). The Word Frequency value provided by Coh-Metrix uses the CELEX word-form frequencies based on the COBUILD/Birmingham corpus of 17.9 million written and spoken word tokens,[1] but since CELEX word frequency is packed in the output of Con-Metrix, we report this value in the study.

Word Familiarity calculates how familiar a word is to an adult. Coh-Metrix measures characteristics of words such as Word Familiarity from the MRC Psycholinguistic Database, a machine-usable dictionary, which collected adult ratings of the content words on a 7-point scale (Coltheart, 1981; Wilson, 1988). Higher scores indicate higher familiarity and reflect easier processing (Graesser et al., 2011).

Word Concreteness refers to how concrete or non-abstract a word is. To calculate word concreteness, Coh-Metrix also uses human ratings collected in the MRC Psycholinguistic Database (Coltheart, 1981; Wilson, 1988). A word that refers to an object, material, or person is, generally, more concrete than an abstract word, and receives a higher concreteness score. If a text has a higher percentage of content words, which are concrete rather than abstract, it receives a higher concreteness score than those with more abstract words.

Word Imageability calculates how easy it is to construct a mental image of the word, measured by adult human ratings in the MRC Psycholinguistic database (Coltheart, 1981; Wilson, 1988). Based on Graesser et al. (2011), the Word Imageability scores provided by Co-Metrix indicate whether a word evokes mental images easily, a higher score on the scale indicating an easier evocation of mental images and possible easy processing.

Finally, ***Hyperonymy*** refers to word specificity regarding the hierarchical relations of words. To report the Word Hyperonymy value for all content words, nouns, and verbs, Coh-Metrix uses a lexical database called WordNet® (Fellbaum, 1998; Miller, Beckwith, Fellbaum, Gross, & Miller, 1990) from Princeton University. Here, we calculated the hypernymy values only for nouns. A higher Hyperonymy score for nouns

1. For a more up-to-date corpus, word frequency can be counted using corpus such as SUBTLEX (Brysbaert & New, 2009).

indicates an overall use of more specific nouns in the text (e.g., *folding chair* and *camp chair* would receive higher hypernymy values than *chair* in the WordNet®).

Analysis

The 26 speech samples, of which 13 were the learners' first speech performance samples, and 13 their second speech performance samples, were transcribed into text documents by one of the researchers and checked by another. In view of the small sample size and the paired sample, data were analyzed using a related-samples Wilcoxon Signed Rank Test, which is a non-parametric test designed to evaluate the difference between two conditions which are correlated (McCrum-Gardner, 2008). We used SPSS 23 for the analyses comparing the 13 participants' first speaking with their second speaking performance. The null hypothesis for the test asserts that the medians of the variable under investigation in the two samples are identical. Significance was assessed relative to an alpha level of .05 ($p < .05$). Effect size r was measured using Cohen's (1992) criteria: an r of 0.1 indicates a small effect, 0.3 indicates a medium effect, and 0.5 indicates a large effect.

Results

The results of descriptive and inferential statistics for discourse performance measures are given in Table 3, using a related-samples Wilcoxon Signed Rank Test to evaluate the difference between the first and the second speech performances of the 13 learners. Asterisks indicate significant results to an alpha level of .05 ($p < .05$). We summarize the results in terms of the three aspects of discourse quantity, cohesion, and lexical quality.

Results for discourse quantity

First, with regard to discourse quantity, the learners produced a significantly larger quantity of discourse in the second performance than in the first, which is shown by all three measures of discourse quantity: number of Total Words ($Z = -3.11$, $p < .01$, $r = .61$), number of Propositions ($Z = -3.11$, $p < .01$, $r = .61$), and P-density ($Z = -2.69$, $p < .01$, $r = .53$), with large effect sizes. The results confirm our hypothesis that task repetition helps increase the overall quantity of the oral narrative discourse, including the total words produced, the number of ideas, and idea density (Kintsch, 1974). This can be explained by the feedback that arises through the 'monitoring' opportunity from the first performance. That is, the macro- and micro-planning of that

Table 3. Analysis results for the discourse performance measures

	Performance						
	1st round		2nd round		*Z*	*p*	*r*
Discourse quantity	Median	Mean (SD)	Median	Mean (SD)			
Total Words	285	336 (95.9)	393	418 (94.2)	−3.11	.002**	0.61
Propositions	124	143 (47.6)	193	190 (49.4)	−3.11	.002**	0.61
P-density	.435	0.423 (0.05)	.452	0.454 (0.04)	−2.69	.007**	0.53
Cohesion							
LSA Given/New	.333	0.329 (0.02)	.349	0.353 (0.02)	−2.59	.01**	0.51
Causal Connectives	20.72	24.17 (11.2)	25.45	28.00 (10.23)	−1.29	.196	0.25
Causal Ratio	0.095	0.10 (0.05)	0.130	0.15 (0.11)	−1.45	.147	0.28
Lexical Quality							
MTLD	35.77	34.82 (5.78)	35.55	34.59 (5.31)	−.245	.807	0.05
vocd-D	44.72	42.15 (7.56)	44.97	43.56 (6.60)	−.245	.807	0.05
Word Frequency	3.220	3.211 (0.075)	3.191	3.209 (0.076)	−.524	.600	0.10
Word Familiarity	585.38	584.9 (4.32)	584.48	585.0 (3.13)	−.175	.861	0.03
Word Concreteness	417.19	415.3 (15.4)	417.32	415.5 (14.9)	−.314	.753	0.06
Word Imageability	429.60	429.7 (14.0)	433.52	430.6 (13.7)	−.314	.753	0.06
Noun Hypernymy	7.27	7.27 (0.22)	7.63	7.50 (0.32)	−2.34	.019*	0.46

*$p < .05$, **$p < .01$

first performance effectively extended into the second performance when the speaker repeated the same task in the same condition. Macro-planning prepares for the narrative ideas, and micro-planning helps generate the discourse forms to express those ideas, both of which are relevant to improving discourse quantity.

Results for cohesion

Second, part of our hypothesis regarding task repetition effects on discourse cohesion was confirmed as well. The results for the LSA Given/New measure show significantly higher LSA scores in the learners' second speech performance compared to their first with a large effect size ($Z = -2.59$, $p < .05$, $r = .51$). However, the results for the two other measures, Causal Connectives and Causal Ratio, were not significant (Causal Connectives: $Z = -1.29$, $p = .196$, $r = 0.25$; Causal Ratio: $Z = -1.45$, $p = .147$, $r = 0.28$). The higher LSA Given/New scores in the learners' second speech performance indicates improved cohesion in the second narrative speech performance, again likely due to the feedback received from 'monitoring' by the speaker's self-comprehension system in the first task enactment, as well as the extended macro- and micro-planning when repeating the task.

Results for lexical quality

Third, with regard to lexical quality, only one of the seven indices measuring the lexical quality of the narrative discourse showed statistically significant findings, the index of Noun Hypernymy value, with a significantly higher score in the second speech performance than in the first and a moderate effect size (Noun Hypernymy: $Z = -2.34$, $p < .05$, $r = .46$). This indicates that, overall, the learners used more specific nouns than general nouns in the second performance. Once again the reasons are probably related to time and attentional space afforded to the L2 speakers by the task repetition condition, so that learners were able to access their mental lexicon to map the lexical forms with their conceived ideas more precisely than by using more general words in the first speech performance. However, the other indices for measuring lexical quality did not show significant results. In fact, the non-significant results of lexical diversity (i.e., MTLD and vocd-D) are generally consistent with the literature findings about the lack of effects of task repetition on lexical performance. However, because they are relatively new in TBLT research, these indices as well as other lexical measures need to be explored in future studies.

To summarize, this study found that immediate task repetition resulted in a greater quantity of spoken discourse (evidenced by the Total Words, Propensity, and P-density scores), higher discourse cohesion (evidenced by LSA Given/New), and the use of more specific nouns regarding lexical quality (evidenced by Noun

Hypernymy value). We turn now to the theoretical as well as pedagogical implications of these results.

Discussion

Discourse performance in TBLT

This study contributes to the literature by exploring changes in discourse performance in task-based learning, in terms of the three aspects of discourse measures: discourse quantity, cohesion, and lexical quality. These three aspects, we believe, are related to the conceptualization stage in Levelt's speech production model and the products of the learners' macro- and micro-planning in conceptual preparation as well as those of 'lexical selection', as learners are mapping the conceived pre-verbal messages with lemmas in the mental lexicon (Levelt, 1989, 1999). The study found evidence of such changes in discourse performance in the immediate task repetition condition. In this condition the speakers are watching a new story unfamiliar to them on the first occasion and narrate what they see on the screen in real time as they watch, so clearly they have to organize their propositions, connect them together, and access vocabulary on the spot without advance preparation, which is of course challenging. In contrast, when they come to repeat the same task, this provides learners opportunities to think ahead of time and prepare their propositions in advance for what they are going to say next. This would explain performance changes in discourse quantity, cohesion, and lexical quality of the speech products. This process is especially meaningful for L2 learners because while they prepare for propositional content, organize cohesion, and search for more appropriate vocabulary on the second occasion, the memory of the first enactment of the task is still available and thus likely to provide a "priming" effect to enhance the speakers' second speech performance.

These discourse performance changes are also consistent with Grice's (1975) cooperative principle (i.e., people can achieve effective communication in conversation by obeying the maxims of quality, quantity, relevance, and manner), although Grice's maxims are often used to observe and explain conversation between interlocutors. In this study, the L2 learners, while they performed the speaking tasks to imagined listeners, took the dual role of speaker and listener. Accordingly, they worked hard to increase the speech quantity, improve cohesion, and use more specific nouns while performing the same narrative task again, in a sense in line with Grice's (1975) maxims of quantity, relevance, and quality. This argument has been made by Widdowson (1980) in that the writer's (in our study the speaker's) "covert and non-reciprocal interaction" with herself enables the communication and generation of ideas (p. 238), resulting in potential improvement in the quality of the resulting written (here the

spoken) discourse. The broader implication is that when researching language tasks, it is important to consider the impact of task design and implementation conditions on aspects of L2 discourse performance, as a language task is fundamentally a context for practicing communicative functions in the real world.

Implications for TBLT pedagogy

The findings of this study have pedagogical implications, supporting Bygate's (1996, 1999, 2001) argument with empirical evidence. For meaningful tasks, repetition will not lead to verbatim repetitive performance, but, rather, to learners re-constructing their interlanguage each time, using the opportunity to frame and reframe their language. In this study, the learners did not merely repeat the same discourse performance of their first task performance. Instead, they increased the discourse quantity, improved the cohesion, and used more specific nouns than before. Therefore, in the classroom, teachers can use simple, immediate task repetition manipulation for enhancing the spoken discourse performance of L2 learners.

Conclusion and future research

This chapter investigated the effects of immediate task repetition on oral discourse performance of L2 learners, using several computer-generated indices to measure three sub-dimensions of discourse performance: discourse quantity, cohesion, and lexical quality. The results show that, in their second speech performance, learners produced more total words, higher proposition scores, and higher P-density values, reflecting an increase in discourse quantity. They also showed higher LSA Given/New values for speech cohesion, and the use of more specific nouns indicated by the higher Noun Hyperonymy values, showing a change in the lexical quality of students' second speech performance. Due to the small number of participants in this study, any generalization of the task repetition effects on L2 speech discourse performance should be tempered with caution. Future research should investigate not only the discourse performance of L2 learners under other task conditions (e.g., +/−pre-task planning) and task characteristics (e.g., monologic/dialogic tasks, see Cognition Hypothesis, Robinson, 2001), but also the discourse measures that reflect both L2 speech processing and the discourse functions of communication.

Acknowledgement

This work was supported by grants from the Pittsburgh Science of Learning Center (NSF #SBE-0836012 & #SBE-0354420) and Hong Kong RGC #27606915. The authors wish to thank Martin Bygate, Peter Skehan, and the series editors John Norris and Kris van den Branden for their help and insightful comments. All errors and omissions are the authors' responsibility alone.

References

Ahmadian, M. J., & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, fluency, and complexity of EFL learners' oral production. *Language Teaching Research*, 15, 35–59. https://doi.org/10.1177/1362168810383329

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (release 2)*. Philadelphia, PA: Linguistic Data Consortium.

Bachman, L. F. 1990. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford: Oxford University Press.

Brown, C., Snodgrass, T., Kemper, S., Herman, R., & Covington, M. (2008). Automatic measurement of propositional idea density from part-of-speech tagging. *Behavior Research Methods*, 40(2), 540–545. https://doi.org/10.3758/BRM.40.2.540

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Bygate, M. (1996). Effects of task repetition: Appraising the developing language of learners. In D. Willis & J. Willis (Eds.), *Challenge and change in language teaching* (pp. 134–146). London: Heinemann.

Bygate, M. (1999). Task as context for the framing, reframing and unframing of language. *System*, 27(1), 33–48. https://doi.org/10.1016/S0346-251X(98)00048-7

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing* (pp. 23–48). Harlow: Longman.

Bygate, M., & Samuda, V. (2005). Integrative planning through the use of task-repetition. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 37–76). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.11.05byg

Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. C. Richard & R. W. Schmidt (Eds.), *Language and Communication* (pp. 2–14). London: Longman.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47. https://doi.org/10.1093/applin/1.1.1

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, 33, 497–505. https://doi.org/10.1080/14640748108400805

Cook, G. (1989). *Discourse. (Language Teaching: A Scheme for Teacher Education)*. Oxford: Oxford University Press.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Crossley, S. A. (2013). Assessing automatic processing of hypernymic relations in first language speakers and advanced second language learners: A semantic priming approach. *The Mental Lexicon*, 8(1), 96–116. https://doi.org/10.1075/ml.8.1.05cro

Crossley, S. A., Cobb, T., & McNamara, D. S. (2013). Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications. *System*, 41(4), 965–981. https://doi.org/10.1016/j.system.2013.08.002

Crossley, S. A., & Salsbury, T. (2010). Using lexical indices to predict produced and not produced words in second language learners. *The Mental Lexicon*, 5(1), 115–147. https://doi.org/10.1075/ml.5.1.05cro

Crossley, S., Salsbury, T., & McNamara, D. (2010). The development of polysemy and frequency use in English second language speakers. *Language Learning*, 60(3), 573–605. https://doi.org/10.1111/j.1467-9922.2010.00568.x

de Jong, N., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, 61, 533–568. https://doi.org/10.1111/j.1467-9922.2010.00620.x

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Ellis, R. (2005). Integrative planning through the use of task-repetition. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 3–36). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.11.03ell

Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity and accuracy in l2 oral production. *Applied Linguistics*, 30, 474–509. https://doi.org/10.1093/applin/amp042

Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 4(2), 139–155. https://doi.org/10.1016/1060-3743(95)90004-7

Fellbaum, C. (1998) *WordNet: An Electronic Lexical Database*. Cambridge, MA: The MIT Press.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21(3), 354–375. https://doi.org/10.1093/applin/21.3.354

Fukuta, J. (2016). Effects of task repetition on learners' attention orientation in L2 oral production. *Language Teaching Research*, 20(3), 321–340.

Gass, S., Mackey, A., Álvarez-Torres, M. J., & Fernández-García, M. (1999). The effects of task repetition on linguistic output. *Language Learning*, 49, 549–581. https://doi.org/10.1111/0023-8333.00102

Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, 2, 371–398. https://doi.org/10.1111/j.1756-8765.2010.01081.x

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40, 223–234. https://doi.org/10.3102/0013189X11413260

Graesser, A., McNamara, D. S., Louwerse, M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers*, 36, 193–202. https://doi.org/10.3758/BF03195564

Grice, P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics, 3: Speech acts* (pp. 41–58). New York, NY: Academic Press.

Hempelmann, C. F., Dufty, D., McCarthy, P. M., Graesser, A. C., Cai, Z., & McNamara, D. S. (2005). Using LSA to automatically identify givenness and newness of noun phrases in written discourse. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), Proceedings of the 27th Annual Conference of the Cognitive Science Society (pp. 941–946). Mahwah, NJ: Lawrence Erlbaum Associates.

Hinkel, E. (2004). *TOEFL test strategies* (3rd ed.). Hauppauge, NY: Barron's Educational Series.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 461–473. https://doi.org/10.1093/applin/amp048

Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning*, 63(supplement 1), 87–106. https://doi.org/10.1111/j.1467-9922.2012.00739.x

Kintsch, W. A. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lambert, C., Kormos, J., & Minn, D. (2017). Task repetition and second language speech processing. *Studies in Second Language Acquisition*, 39(1), 167–196.

Landauer, T., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

Levelt, W. J. M. (1999). Producing spoken language: A blueprint of the speaker. In C. Brown & P. Hagoort (Eds.), *Neurocognition of language* (pp. 83–122). Oxford: Oxford University Press.

Levelt, W. J. M., & Indefrey, P. (2000). The speaking mind/brain: Where do spoken words come from? In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain papers from the First Mind Articulation Project Symposium* (pp. 77–93). Cambridge, MA: The MIT Press.

Long, M. (2014). *Second language acquisition and task-based language teaching*. Chichester: John Wiley & Sons.

Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014. *International Journal of Corpus Linguistics*, 22(3), 319–344.

Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, 4, 221–50.

Malvern, D. D., Richards, B. J., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Houndmills: Palgrave Macmillan. https://doi.org/10.1057/9780230511804

McCarthy, P. M. (2005). *An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD)* (Unpublished doctoral dissertation). University of Memphis, Memphis, TN.

McCarthy, M., & Carter, R. (2014). *Language as discourse: Perspectives for language teaching*. London: Routledge.

McCarthy, P. M., Dufty, D., Hempelman, C., Cai, Z., Graesser, A. C., & McNamara, D. S. (2012). Newness and givenness of information: Automated identification in written discourse. In P. M. McCarthy & C. Boonthum-Denecke (Eds.), *Applied natural language processing and content analysis: Identification, investigation, and resolution* (pp. 457–478). Hershey, PA: IGI Global.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2), 381–392. https://doi.org/10.3758/BRM.42.2.381

McCrum-Gardner, E. (2008). Which is the correct statistical test to use? *British Journal of Oral and Maxillofacial Surgery*, 46(1), 38–41. https://doi.org/10.1016/j.bjoms.2007.09.002

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511894664

McNamara, D. S., Louwerse, M. M., Cai, Z., & Graesser, A. (2005, January 1). Coh-Metrix version 1.4. Retrieved from <http//:cohmetrix.memphis.edu> (2 July, 2016).

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes*, 47, 292–330. https://doi.org/10.1080/01638530902959943

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1990). *Five papers on WordNet* (Technical Report No. 43). Princeton, NJ: Cognitive Science Laboratory, Princeton University. Retrieved from <http://wordnetcode.princeton.edu/5papers.pdf>

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in SLA: The case of complexity. *Applied Linguistics*, 30(4), 555–578. https://doi.org/10.1093/applin/amp044

Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge: Cambridge University Press.

Qiu, X., & Lo, Y. Y. (2017). Content familiarity, task repetition and Chinese EFL learners' engagement in second language use. *Language Teaching Research*, 21(6), 681–698.

Robinson, P. (2011). Task-based language learning: A review of issues. *Language Learning*, 61(Supplement 1), 1–36. https://doi.org/10.1111/j.1467-9922.2011.00641.x

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. Houndmills: Palgrave.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency and lexis. *Applied Linguistics*, 30(4), 510–532. https://doi.org/10.1093/applin/amp047

Skehan, P., Bei, X., Li, Q., & Wang, Z. (2012). The task is not enough: Processing approaches to task-based performance. *Language Teaching Research*, 16(3), 170–187. https://doi.org/10.1177/1362168811428414

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49(1), 93–120. https://doi.org/10.1111/1467-9922.00071

Templin, M. (1957). *Certain language skills in children*. Minneapolis MN: University of Minnesota Press.

Van den Branden, K., Bygate M., & Norris, J. (Eds.) (2009). *Task based language teaching: A reader*. Amsterdam: John Benjamins.

Wang, Z. (2014). On-line time pressure manipulations: L2 speaking performance under five types of planning and repetition conditions. In P. Skehan (Eds.), *Processing Perspectives on Task Performance* (pp. 27–62). Amsterdam: John Benjamins.

Widdowson, H. G. (1978). *Teaching language as communication*. Oxford: Oxford University Press.

Widdowson, H. G. (1980). Conceptual and communicative functions in written discourse. *Applied linguistics*, 1, 234. https://doi.org/10.1093/applin/1.3.234

Wilson, M. D. (1988). The MRC psycholinguistic database: Machine readable dictionary, version 2. *Behavior Research Methods, Instruments, and Computers*, 20(1), 6–11. https://doi.org/10.3758/BF03202594

CHAPTER 5

The impact of intra-cultural and inter-cultural task repetition on interaction

Scott Aubrey
Kansai University

Previous classroom-based research has shown that task repetition (Kim, 2013) and interlocutor proficiency (e.g. Watanabe & Swain, 2007) can affect the extent to which learners reflect on language form during the performance of collaborative tasks. Furthering these lines of inquiry, this study compares the effect of task repetition on Japanese EFL learners' attention to linguistic form during the performances of oral tasks under two conditions: when task repetition involves (1) a Japanese interlocutor and (2) a non-Japanese interlocutor. Japanese learners (n = 36) from two EFL classes performed two oral, collaborative tasks in pairs. In one class, learners were paired with international, English-speaking students (n = 18), while in the other class learners were paired with peers of the same class. Participants in each group kept the same partner for the duration of the initial and repeated task performances. Task interactions were transcribed and analyzed in terms of language-related episodes (LREs). Results indicated that repeating tasks with a non-Japanese interlocutor led to a number of significant benefits, including a higher incidence of LREs overall, more episodes being resolved with uptake, and a higher rate of complex and grammatical LREs than during the first task performances. On the other hand, repetition with a Japanese interlocutor resulted in no significant change in LRE production, but a significant decrease in the production of lexical LREs.

Introduction

Research in the field of second language acquisition has shown that tasks can provide both opportunities for meaning-focused communication and opportunities for learners to attend to form during the interactions that arise (e.g. Leeser, 2004; Watanabe & Swain, 2007; Williams, 2001). In addition, the ways in which the same tasks are repeated can give rise to changes in task performance (e.g. Kim, 2013). Task repetition is one implementation variable that not only has a marked effect on task performance but is itself multifaceted, encompassing variations that involve changes in task content

DOI 10.1075/tblt.11.05aub

and procedure (Pattansorn, 2010). However, the repeated performance of a task rarely occurs under the exact same conditions. Further obscuring the distinction between types of task repetition is the fact that collaborative tasks can – and often are – repeated in the classroom with different interlocutors.

Despite the increasing exploration of different types of task repetition, the intertwined nature of task repetition and interlocutor variation is rarely acknowledged in the literature. There appear to be few, if any, studies that have systematically looked at how working with interlocutors of different cultural backgrounds impacts task repetition. In an attempt to understand how inter-cultural contact affects task engagement, Aubrey (2017) demonstrated that learners who performed tasks with a culturally unfamiliar interlocutor were more emotionally engaged (i.e., reported higher levels of 'flow') and generated more interactive performances (i.e., engaged in a greater amount of talk) than learners who performed the same tasks with a participant from their own culture. Although the effect of task repetition was not investigated in that study, it was an important step forward in terms of exploring the potential benefits of inter-cultural interaction in the EFL classroom. The current study, then, is a follow-up to Aubrey (2017), expanding on the dataset to include transcripts of repeated task performances, which were analyzed to reveal the effect of intra-cultural and inter-cultural task repetition on learners' attention to linguistic form. In doing so, this study provides an example of how inter-cultural task-based interaction might be implemented in the EFL classroom.

Attention to form during task interaction

Previous research on the effects of task repetition has had various aims. In regards to how task repetition has impacted performance, studies have looked at features of oral production (e.g., Bygate, 1996; Kim & Tracy-Ventura, 2013; Patanasorn, 2010) and the complexity, accuracy, and fluency of oral production (e.g. Ahmadian & Tavakoli, 2011; Bygate, 2001; Kim & Tracy-Ventura, 2013). A less common approach is to look at the degree to which learners pay attention to form during task performances and how this is affected by subsequent performances of the same or similar tasks.

As Manchon (2014) states: "the rationale for making task repetition central in TBLT preoccupations is ultimately related to the general concern of finding ways of promoting attention to language in TBLT-based/oriented pedagogical approaches" (p. 18). Skehan (1998) suggests that learners' attentional or processing capacity is limited, giving rise to a competition between meaning and form. Similar to effects of pre-task planning (e.g., Skehan & Foster, 1999), learners may benefit from repeating a task by having already conceptualized the message content, allowing them to more easily devote their attention to form.

One way in which learners could do so is through interactional focus on form, which refers to instances when learners attend to "linguistic elements as they arise incidentally in lessons whose overriding focus is on meaning, or communication" (Long, 1991, p. 46). The Output Hypothesis (Swain, 1985, 1995, 2005) formed the basis for a concept Swain has referred to as 'collaborative dialogue' (Swain, 2000): This is a kind of focus on form that occurs when learners attempt to resolve linguistic problems through interaction. Swain (2006) describes these interactional episodes as a cognitive and social activity that leads to the construction and shaping of linguistic knowledge. Furthermore, it affords learners the opportunity to modify their output, which can lead to positive developmental effects (Swain, 1995). Instances of collaborative dialogue have been operationalized as language-related episodes (LREs) or "any part of a dialogue where the students talk about the language they are producing, question their language use, or correct themselves" (Swain & Lapkin, 1998, p. 328).

In addition to task repetition, interlocutor proficiency has been found to impact how learners' attend to form during task interaction (e.g., Leeser, 2004; Kim & McDonough, 2008). Changing interlocutors when a task is repeated means that learners have access to a different source of input. How an interlocutor's input facilitates attention to form depends on a variety of factors, which may include proficiency differences between the interlocutors and the familiarity of their interlocutor (and hence willingness to enter into collaborative dialogue). A review of studies relating to these two implementation factors follows below.

Effect of task repetition on LREs

One study that used the LRE construct to examine the effect of repetition on interaction is that by Kim (2013). In framing her study, she drew heavily on the research of Mackey, Kanganas, and Oliver (2007), who examined the interactional feedback between young learners engaging in communicative tasks that were familiar or unfamiliar to them either in terms of procedure or content. The three main findings of Mackey et al. were:

- Learners who performed unfamiliar tasks produced more clarification requests and confirmation checks, as well as more corrective feedback in general, than those who performed familiar tasks.
- Learners who engaged in tasks under the procedurally familiar condition had more opportunities to use feedback than those performing unfamiliar tasks.
- Learners who engaged in tasks that were familiar were more likely to use feedback than learners performing unfamiliar tasks.

Based on these results, Mackey et al. suggested that the effects of the two types of familiar tasks (content and procedure) related more to improved quality of task performance (i.e., increased fluency), whereas the performance of unfamiliar tasks encouraged more attention to form through corrective feedback.

Following up on these findings, Kim (2013) focused on the comparison of the impact of task repetition and procedural repetition on Korean EFL learners' production of LREs. Two intact EFL classes were randomly assigned to carry out three oral tasks with either task repetition or procedural repetition. The task repetition group performed three of the same tasks (same content and procedure) over a one-week period, whereas the procedural repetition group carried out three similar tasks (same procedure, different content). LREs were identified in the transcripts and categorized in terms of linguistic focus and resolution. Kim found that the procedural repetition group produced significantly more lexical LREs and grammatical LREs than the task repetition group. In fact, while this group had fewer LREs each time they repeated the task, the procedural repetition group generated a larger number of LREs for each performance. However, similar to Mackey et al.'s (2007) finding, the task repetition group was able to resolve more of their LREs correctly, suggesting that learners divert some of their attention to the activity of resolution when tasks are repeated.

In addition to analyzing LRE quantity and quality, Kim (2013) had students complete a task perception questionnaire and conducted interviews with students after the final task performance, which shed light on their attitudes towards the repetition types. She found that students in the task repetition group "tended to lose motivation easily" as they became "tired of doing the same thing" (p. 17). Thus, the likelihood of generating LREs may have been negatively affected by a suppressed willingness to communicate during repeated performances.

Effect of interlocutor on LREs

In both second language classrooms and second language studies, it is often the case that learners repeat a task with a different interlocutor. For example, when Lynch and Maclean (2000) investigated the repetition effects of a 'poster carousel', the repeated task performances involved answering questions from different participants. The characteristics of each new interlocutor, such as their level of familiarity with the presenter or their linguistic proficiency, as well as the content of the questions that were asked, affected the learners' repeated performances.

In terms of proficiency differences, several studies have shown that interacting with a more proficient peer leads to the production of more LREs (e.g. Bowles, Adams, & Toth, 2014; Kim & McDonough, 2008; Leeser, 2004). For example, Leeser (2004) investigated how grouping learners by their relative proficiency affected

collaborative dialogue during a dictogloss task completed by Spanish L2 learner pairs. He found that if there was at least one high proficiency interlocutor, significantly more LREs, both lexical and grammatical, were generated than with low-proficiency pairs. Furthermore, higher-proficiency and mixed-proficiency pairs were much more likely than low-proficiency pairs to produce LREs that focus on grammatical as opposed to lexical items. Leeser speculated that the difference in LRE production could be attributed to the relative difference in task demands: "it could be the case that the lower proficiency learners were struggling just to extract meaning from the passage" (p. 73). While lower-proficiency dyads focused on those elements that carried the most meaning (i.e., produced a similar proportion of grammatical and lexical LREs), for higher-proficiency dyads, the task demand was much less, so they could devote more of their attention to grammatical form (i.e. produce more grammatical LREs).

In a similar line of inquiry, Kim and McDonough (2008) examined the occurrence and resolution of LREs when Korean L2 learners collaborate with interlocutors of different and similar proficiency levels. Interaction was generated via dictogloss tasks, with all learners performing the tasks with both an intermediate and advanced proficiency interlocutor. Results showed that, in terms of LRE production, collaborative dialogue with advanced interlocutors contained significantly more lexical LREs and correctly resolved LREs, which supports the findings of previous studies (Leeser, 2004; Watanabe & Swain, 2007; Williams, 2001). However, unlike previous findings, there was no significant difference between the two groups in terms of grammatical LREs. In fact, in contrast to Leeser's (2004) findings, Kim and McDonough found that interaction with a higher proficiency learner led to a lower proportion of grammatical LREs than interaction between lower proficiency learners.

In addition to proficiency, other studies have looked at the effect of interlocutors with different proficiency and cultural affiliation. For example, Fernández Dobao (2012) examined the effect of a native speaker interlocutor, comparing the task interactions of learner-learner and learner-native speaker pairs. Twenty-four EFL learners and eight native English speakers were paired according to four different dyad kinds: (1) learner (advanced)-learner (advanced), (2) learner (intermediate)-learner (intermediate), (3) learner (advanced)-native speaker, and (4) learner (intermediate)-native speaker. As with studies looking at high-proficiency interlocutors (e.g. Leeser, 2004; Kim and McDonough, 2008), it was found that learner-native speaker interaction resulted in significantly more LREs produced than learner-learner interaction. Fernández Dobao observed that learners who were paired together often avoided entering into LREs when they had the chance to because "they lacked the linguistic resources needed to collaborate with their peers in the building of new lexical knowledge" (p. 23). Learners tended to prefer successful communication of the message to a focus on linguistic accuracy.

The above studies demonstrate that collaborative dialogue is, in general, more productive in terms of LRE generation when tasks are performed with at least one interlocutor of high proficiency. Furthermore, challenging task demands may constrain lower proficiency learners to devote their attention to meaning rather than to the resolution of grammatical problems. Task repetition may serve to lessen this cognitive burden, allowing learners to focus more on form-related features of language during the repeated performance.

The current study

As a follow-up study to Aubrey (2017), the aim of this study is to explore the impact on interaction of inter-cultural and intra-cultural task repetition in an EFL classroom. To examine the interactional features of task performances, this study used the framework of language-related episodes (LREs). In an effort to provide a fine-grained analysis, it adopts an expanded framework, which takes into account the type, linguistic focus, complexity, and uptake of such episodes. Specifically, this chapter attempts to answer the following research questions:

1. How does intra-cultural task repetition affect the frequency, type, linguistic focus, complexity, and resolution of language-related episodes generated during oral, collaborative tasks?
2. How does inter-cultural task repetition affect the frequency, type, linguistic focus, complexity, and resolution of language-related episodes during oral, collaborative tasks?

For the purpose of this study, 'intra-cultural task repetition' refers to a learner repeating a task with an interlocutor who is a member of the *same* cultural and linguistic community, whereas 'inter-cultural task repetition' occurs when the task participants are from *different* cultural and linguistic communities. It is also worth noting that the inter-cultural dimension as implemented in this study implicates a difference in proficiency while the intra-cultural dimension occurs between similar proficiency learners.

Method

Participants

A total of 63 students studying at a private university in Japan participated in this research. Forty-two of these participants were first-year (domestic) Japanese students,

beginning their first semester in a one-year intensive English program. The participants were drawn from two intact classes consisting of 21 students each, both of which were taught by the same teacher. Participants in each class majored in various subjects, including Theology, Economics, Sociology, and Business. In order to gain entry to the intensive program, students were required to have a paper-based TOEFL score of between 430 and 500. Thus, all participants could be described as having intermediate-level English proficiency. The two intact classes were randomly assigned to the Inter-cultural and Intra-cultural group conditions – the label corresponding to the kind of interlocutor students in each group were paired with.

The remaining twenty-one participants were international, non-Japanese students, who were invited into one of the aforementioned EFL classes to create the inter-cultural condition. The participants were selected by the researcher based on the following criteria: (1) not born in Japan, (2) either a native English speaker or in possession of 'very good' English-speaking skills, (3) committed to attending all classes involving the research activities, and (4) interested in inter-cultural exchanges. Of the 21 international students who participated in the study, 17 students were short-term study abroad students visiting the university for either one or two semesters while four students were aiming to obtain a degree, and therefore intended to stay at the university for the duration of their program. A background questionnaire revealed that thirteen of the international students identified themselves as native English speakers, who reported having either American ($n = 10$), Singaporean ($n = 1$), or Australian ($n = 2$) nationality. The remaining eight non-native English speakers identified themselves as Mexican ($n = 1$), German ($n = 2$), Lithuanian ($n = 1$), Indonesian ($n = 1$), Norwegian ($n = 1$), French ($n = 1$), and Canadian ($n = 1$). The eight non-native English-speaking students self-reported speaking the following languages as an L1: French ($n = 2$), Norwegian ($n = 1$), German ($n = 2$), Spanish ($n = 1$), Indonesian ($n = 1$), and Lithuanian ($n = 1$). All eight non-native speakers rated their own English proficiency on a 7-point Likert scale anchored by 1 (very poor) and 7 (very good). Their means suggested their proficiency levels were 'very good' for reading ($M = 6.75$, $SD = .71$), writing ($M = 6.75$, $SD = .71$), listening ($M = 6.63$, $SD = .52$), and speaking ($M = 6.75$, $SD = .46$).

Research design

The overview of the design of the study is shown in Table 1.

The study took place over a period of 11 weeks. The research period was divided into two stages of five weeks each, with the first stage comprising the initial task performances and the second stage comprising the repeated task performances. During the first stage, five oral tasks were performed – one each week for five weeks with a one-week break in-between. The five tasks were then repeated during the second stage

Table 1. Design of the study

Inter-cultural task repetition	Intra-cultural task repetition	
↓	↓	
Inter-cultural group (N = 21 Japanese students) *9 dyads, 1 triad	Intra-cultural group (N = 21 Japanese students) *9 dyads, 1 triad	Initial task performances
↓	↓	
Task 1	Task 1	Week 1
Task 4	Task 4	Week 4
Inter-cultural group (N = 21 Japanese students) + 21 International students *21 dyads	Intra-cultural group (N = 21 Japanese students) *9 dyads, 1 triad	Repeated task performances
↓	↓	
Task 1R	Task 1R	Week 7
Task 4R	Task 4R	Week 10

with a new interlocutor. The six-week gap between performances of the same task is similar in duration to Bygate (2001). Because of absences during Weeks 2, 3, and 5, it was decided only to use data from classes in which all students were present (i.e. data from Task 1 and Task 4) – this is reflected in Table 1.

For the initial task performances, Japanese students were paired with another student from the same group; that is, nine dyads and one triad in each group performed the tasks (Task 1, Task 4). Students kept the same partner for each task. During the repeated task performances, the same tasks were repeated (Task 1R, Task 4R) under the two above-mentioned conditions. Each Japanese student from the Inter-cultural group performed the tasks with an international student (21 dyads) while learners in the Intra-cultural group repeated the tasks with a different Japanese peer from the same group (9 dyads, 1 triad). In sum, Task 1 and 4 were each repeated once with a six-week time gap, and the repeated task performances involved a change in interlocutors for both groups. Because of the potential interactional differences between dyads and triads, the group of three was eliminated from the study. Therefore, data from 18 Japanese student participants in each group were included in the analysis.

Tasks

The oral tasks used in this study were designed with the purpose of promoting substantial interaction between students. Each task contained a two-way information exchange followed by a decision-making component. A unique feature of these tasks was that students generated the task content through a *guided research* pre-task, in which they gathered information that was later used during the task performance. Two days before students performed each task, they completed a task research worksheet, where they chose an item related to the task (e.g., a person, place, animal, job) and wrote down information in prescribed categories. Table 2 describes the features of each task.

Table 2. Task descriptions

Task	Topic	Guided research: Research…	Information gap: Exchange information on…	Decision-making: Decide…
Task 1	Travel	A place you would like to travel to.	Local language, famous sites, cost of living, food, history.	Which destination to travel to and activities to do.
Task 4	Careers	A job you are interested in.	Company, duties, responsibilities, qualifications, salary	Which job is more interesting and the steps needed to get the job

The tasks were performed in the classroom during regularly scheduled classes. Students were given 25 minutes to complete each task. All task performances in both the Inter-cultural group and Intra-cultural group were audio-recorded.

Analysis

Identification of LREs

The audio recordings of the initial task performances (Task 1, Task 4) and the repeated task performances (Task 1R, Task 4R) were transcribed for all pairs. Transcripts were prepared for the first 20 minutes of each task performance, which were then used to identify and classify LREs.

Swain and Lapkin (1998) define an LRE as "any part of a dialogue where the students talk about the language they are producing, question their language use, or correct themselves" (p. 328). However, in line with other analytical frameworks (e.g. Ellis, Basturkmen, Loewen, 2001; Fortune & Thorpe, 2001; Loewen, 2005), it was decided that LREs should constitute "identifiable units of collaborative activity" (Fortune &

Thorpe, 2001, p. 46). As such, any episode of self-correction that was initiated and resolved within a single turn was excluded from the LRE count. This avoided the practical problem of distinguishing between false starts – and other disfluent speech behavior – and LREs. In order to understand in more detail the nature of the different LREs, their motivation, and their potential contribution to learning, each LRE was coded in terms of four characteristics: a) linguistic focus, b) type, c), complexity, and b) resolution.

Linguistic focus

Following Storch and Aldosari (2013), LREs were divided into lexical, grammatical, and spelling/pronunciation LREs. Despite an effort to follow this classification system, in the data, the focus of some of the LREs did not clearly fit these categories, such as the issue of converting numerals into spoken English. Therefore, a fourth category, *other* LREs, was created. Example (1) illustrates a grammatical LRE generated in an inter-cultural dyad.

Example 1. Grammatical LRE (Inter-cultural group, Task 4R)

```
197  Etsuko:   And if successful, many money.
198  Kate:     Yes.
199  Etsuko:   Yes… yes. Okay and=
200  Kate:     =We like much money
201  Etsuko:   Yes, much money.
202  Kate:     I would like some
203  Etsuko:   Yes. Much money.
```

Etsuko incorrectly uses the indefinite pronoun "many" before an uncountable noun. As can be seen, Kate recasts this incorrect utterance at line 201 and again at line 203.

Type

The LRE category of *type* refers to episodes that are either reactive or preemptive. Based on Ellis et al. (2001), this categorization is concerned with which of the interlocutors initiated the LRE. Reactive LREs occur when a participant responds to an utterance that is problematic either because the meaning is not clear or because it contains a lexical, grammatical, pronunciation or a spelling error. Preemptive LREs, in contrast, are episodes in which a participant initiates a change in topic (topicalizes) to focus on a language issue, not because of a response to a problematic utterance, but because one student predicts that the feature may be problematic to his/her interlocutor or asks a question about form unrelated to the proceeding utterance.

In Example (2), Kazu initiates a preemptive LRE by predicting that Sandra may not understand the term "monster parents", and so defines the phrase in line 257.

Example 2. Preemptive LRE (Inter-cultural group, Task 2R)

```
253  Kazu:     Disadvantages.
254  Sandra:   Hm… okay, so disadvantages.
255  Kazu:     Hm… do you know monster parents?
256  Sandra:   Uh
257  Kazu:     Students' parents are scold the- the teacher.
258  Sandra:   Yes
```

Complexity

In an attempt to capture the complexity of episodes in the data, LREs were classified into two categories: simple and complex. The importance of measuring the complexity of LREs has been mentioned in previous research. For example, Fortune and Thorpe (2001) stated that, "a numerical count does not… capture the fact that episode length varies from extremely long to extremely short" (p. 152). Whereas simple LREs are episodes in which a language issue is resolved in a single turn, complex episodes involve more than one turn. Ellis et al. (2001) found that complex episodes were significantly more likely to be taken up by the learner. Example (3) is an example of a reactive, lexical LRE that has been classified as *simple*. The episode begins when Haru signals that she does not understand the meaning of "creativity" (line 67). Masa provides a simple definition (line 68), which appears to sufficiently resolve the issue.

Example 3. Simple LRE (Intra-cultural group, Task 4)

```
66  Masa:   Ah nothing… nothing. But he have to have creativity
67  Haru:   Creativity?
68  Masa:   Mean imagination.
69  Haru:   Yes
```

Resolution

In previous studies, the resolution of LREs has almost exclusively been based on Swain (1998) and Leeser (2004), who proposed the categories of "correctly resolved", "unresolved", and "incorrectly resolved." According to Leeser (2004), a correctly resolved episode is one in which a "problem or question was solved correctly either by one learner's self-correction or by one learner answering or correcting the other" (p. 65). A problem with this definition is that it sheds no light on whether a correctly resolved LRE contains evidence of successful uptake. Although some studies have included in their analyses the degree to which problems are successfully resolved through

learner-learner (Yule & Macdonald, 1990) and teacher-learner interaction (Ellis et al., 2001), the occurrence of uptake has not been operationalized within an LRE framework. This study, therefore, will draw on Ellis et al.'s (2001) definition of uptake as an interactional move that is optional, occurs during interactional episodes in which there is a gap in knowledge, and is in reaction to a preceding move in which another participant provides linguistic information. Ellis et al. also distinguished between two kinds of uptake: successful and unsuccessful. Unsuccessful uptake, occurs when the responding student move indicates that there was no attempt to repair, the repair failed, or it failed to acknowledge understanding of the target feature. Successful uptake, on the other hand, is a move in which the student response demonstrates the correct use of the targeted item. Using the above interpretations of uptake, the coding of resolution will be extended to four categories: resolved (with uptake), resolved (without uptake), incorrectly resolved, and unresolved. In Example (4), successful uptake of the item "librarian" is evidenced by the correct production of the word in line 62.

Example 4. Resolved (with uptake) LRE (Inter-cultural group, Task 2R)

```
60  Rika:   Lib… lib
61  Mika:   Librarian.
62  Rika:   Librarian… what job does librarian do?
```

As illustrated above, LREs resolved with uptake were defined as episodes in which a language problem is correctly solved or corrected by one learner and the other learner successfully demonstrates uptake (i.e. subsequently uses the corrected item). Conversely, resolved (without uptake) LREs are episodes in which one learner correctly solves a language problem, but there is no evidence of uptake by his/her interlocutor. Example (5) shows how a grammatical LRE is resolved with a recast (line 107), but fails to be uptaken in the subsequent turns.

Example 5. Resolved (without uptake) (Inter-cultural group, Task 2)

```
102  Kazu:   But some [salary]
103  Junko:           [Okay  ]
104  Kazu:   Is really high.
105  Junko:  Okay, okay. Than seven million?
106  Kazu:   More high.
107  Junko:  More- more high- higher… or some are higher than seven
             million. Okay, what is your choice… tell me=
108  Kazu:   =Vacation.
```

As can be seen above, Kazu does not use Junko's feedback to produce the correct form. Uptake moves in which a learner simply acknowledges feedback (e.g. *yes*, *okay*, *yeah, ah*) were not considered to constitute evidence of uptake (classified as unsuccessful uptake in Ellis et al., 2001).

Inter-coder reliability

After the framework for LRE identification was established and definitions for each category were finalized, the researcher and a research assistant coded 10% of the data (9 transcripts) independently. The research assistant was an Assistant Professor of English who worked at the same university as the researcher. For what constituted an LRE, there was 85% agreement between the two raters. Of the LREs that both raters agreed upon, the following inter-coder reliability values were computed.

Table 3. Inter-coder reliability (Cohen's Kappa) for each category of LRE

	Linguistic focus	Type	Complexity	Resolution
Cohen's Kappa	.949	.820	.905	.751

The inter-coder reliability results for LRE categorization are comparable to other studies that have identified LREs (e.g. Fernández Dobao, 2012; Kim & McDonough, 2008; Bowles et al., 2014). Nevertheless, the reliability of resolution was considerably lower than other categories. Ellis et al. (2001), who identified successful uptake with a Kappa of .82, offers a point of comparison. Based on established benchmarks set by Fleiss (1981) (excellent if Kappa > 0.75), Landis and Koch (1977) (substantial if *Kappa* = .61–.80, almost perfect if *Kappa* = .81–1.00), and Altman (1991) (good if *Kappa* = .61–.80, very good if *Kappa* = .81–1.00), the inter-coder reliability results were considered high enough for the researcher to code the remaining data alone.

Examining task repetition

Analyses were carried out for each group separately to answer each research question. Following the procedures of previous research on paired collaboration, the analysis was carried out on each pair, not on individual students. In other words, as each LRE was generated collaboratively, LRE counts were attributed to pairs not individual students. Table 4 shows the dependent variables for each category of LRE.

Table 4. Dependent variables for each category of LRE

Category of LRE	DV1	DV2	DV3	DV4
Linguistic focus	Lexical	Grammatical	Sp./Pron.	Other
Type	Preemptive	Reactive	–	–
Complexity	Simple	Complex	–	–
Resolution	With uptake	Without uptake	Incorrect	Unresolved

DV = dependent variable; *Sp./Pron.* = spelling and pronunciation

The main objective was to determine if there was a significant difference in LRE production between the initial and the repeated task performances for each group. To do this, two one-way within-subjects MANOVAs were conducted to test for the effect of task repetition in each group. Results for the main effect of task repetition were calculated to determine if there were significant differences in the production of LREs between task performances (i.e. initial and repeated task performances). Results for the effect of group on each dependent variable were calculated to determine if there were significant differences in the production of each LRE subcategory.

Results

Total LREs

Table 5 gives the descriptive statistics for LREs produced for each pair during the initial and repeated task performances.

Table 5. Descriptive statistics for LREs

Group	Inter-cultural group				Intra-cultural group			
	Initial task performances		Repeated task performances		Initial task performances		Repeated task performances	
Task	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Task 1	11.22	6.12	16.22	4.95	10.33	4.15	8.44	3.97
Task 2	8.44	5.27	15.00	6.62	7.77	3.63	5.89	2.67
Total LRE	19.67	8.08	31.22	8.27	18.10	5.51	14.33	4.78

The results from Table 5 are represented visually in Figure 1. A sharp increase can be seen in the mean number of LREs produced by the Inter-cultural group across task performances. In contrast, there is a slight decrease in the number of LREs generated by the Intra-cultural group.

A one-way MANOVA was conducted for both groups separately to test for the effect of repetition. The results indicated a significant main effect for repetition for the Inter-cultural group. That is, inter-cultural repetition resulted in a significant increase in the total frequency of LREs, $F(2,24) = 4.098$, $p = .041$, $d = 1.462$. Furthermore, tests of between-subjects effects revealed a significant increase in LREs from the first to second task performance for Task 1, $F(1,25) = 5.233$, $p = .031$, $d = .971$, and for Task 2, $F(1,25) = 6.658$, $p = .016$, $d = .1.096$. In contrast, there were no significant effects for intra-cultural repetition on LRE production, $F(2,15) = .777$, $p = .477$, $d = .775$.

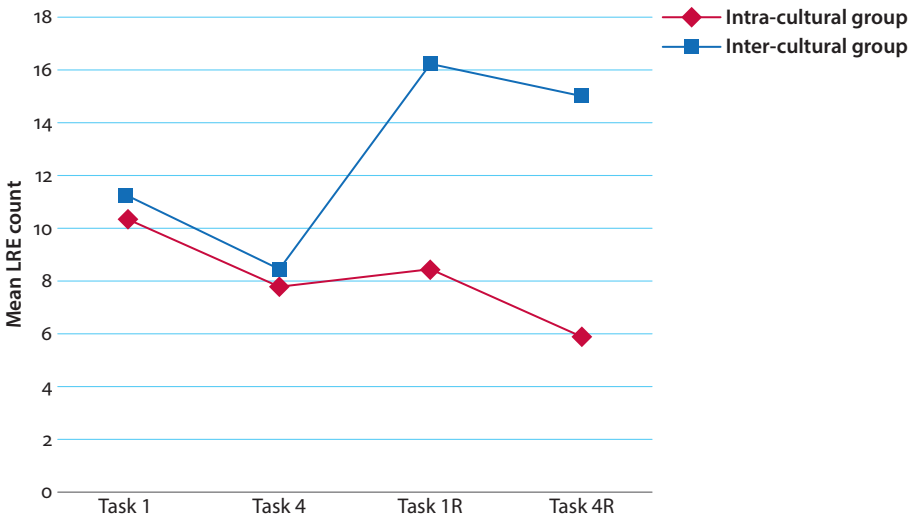

Figure 1. Mean LRE per in each group

Linguistic focus

Table 6 shows the descriptive statistics for the linguistic focus of LREs produced by each group during initial and repeated task performances.

Table 6. Descriptive statistics for LREs (linguistic focus)

	Initial task performances			Repeated task performances		
LRE focus	***M***	***SD***	***%***	***M***	***SD***	***%***
Inter-cultural group						
Lexical	10.22	5.99	51.96	13.55	4.56	43.40
Grammatical	2.56	1.37	13.01	6.17	3.15	19.76
Sp. / Pron.	5.56	4.98	28.27	10.00	4.16	32.03
Other	1.22	.78	6.20	1.5	1.01	4.80
Total LREs	19.67	8.08	100	31.22	8.27	100
Intra-cultural group						
Lexical	10.11	3.09	55.86	7.11	2.70	49.62
Grammatical	2.11	1.67	11.66	1.00	1.03	6.98
Sp. / Pron.	5.11	3.39	28.23	5.33	4.63	37.19
Other	1.78	.78	9.83	.89	1.03	6.21
Total LREs	18.10	5.51	100	14.33	4.78	100

Note: *Sp. / Pron.* = Spelling / Pronunciation

Table 6 shows that pairs in both groups produced more lexical LREs than for any other linguistic focus, comprising over 50% of LREs during the initial task performances, and remaining above 40% in the repeated task performances. In this study, lexical LREs focused on negotiations of word meanings and word choices. This result, therefore, suggests that interactions were mainly prompted by a focus on the meaning of utterances as opposed to form (grammar or spelling/pronunciation).

To investigate changes in the four dependent variables due to task repetition, the results for the multivariate test of between-subjects effects were calculated. These are presented in Table 7.

Table 7. Test of between-subjects effects for repetition

Variable	*F*	*df*	*Sig.*	Direction
Inter-cultural group				
Lexical	2.572	1	.115	
Grammar	9.670	1	.003	RTP > ITP
Sp. / Pron.	5.302	1	.025	RTP > ITP
Other	.445	1	.507	
Intra-cultural group				
Lexical	4.489	1	.042	ITP > RTP
Grammar	2.522	1	.122	
Sp. / Pron.	.018	1	.895	
Other	.064	1	.802	

Note: *Sp. / Pron.* = Spelling / Pronunciation, *ITP* = Initial task performances, *RTP* = Repeated task performances

Results from Table 7 indicate that for the Inter-cultural group there was a significant effect for repetition on the production of grammatical LREs, $F(1,52) = 9.670$, $p = .003$, $d = .915$, and spelling / pronunciation LREs, $F(1,52) = 5.302$, $p = .025$, $d = .677$. Intra-cultural repetition, however, resulted in a significant decrease in the production of lexical LREs, $F(1,34) = 4.489$, $p = .042$, $d = .727$ with all other dependent variables showing no significant change.

Type

Table 8 shows the total LRE frequency counts, broken down by type (reactive and preemptive), generated by each pair for initial and repeated task performances.

Table 8. Descriptive statistics for LREs (type)

	Initial task performance			Repeated task performance		
LRE type	***M***	***SD***	***%***	***M***	***SD***	***%***
Inter-cultural group						
Reactive	14.22	5.60	72.29	24.44	7.36	78.28
Preemptive	5.44	4.32	27.66	6.78	3.12	21.72
Total LREs	19.67	7.95	100	31.22	7.01	100
Intra-cultural group						
Reactive	15.22	4.12	84.04	11.89	4.64	82.97
Preemptive	2.89	2.65	15.96	2.44	1.69	17.03
Total LREs	18.11	4.94	100	14.33	4.63	100

Table 8 reveals that the majority of LREs generated (72–84%) within each group were reactive. In other words, most LREs were produced in response to a misunderstood or problematic utterance.

A one-way MANOVA was conducted for both groups separately to test for the effect of repetition. The effects of repetition on each LRE type are shown in Table 9.

Table 9. Test of between-subjects effects for repetition

Variable	***F***	***df***	***Sig.***	**Direction**
Inter-cultural group				
Reactive	13.909	1	.0005	RTP > ITP
Preemptive	.788	1	.379	
Intra-cultural group				
Reactive	2.434	1	.128	
Preemptive	.173	1	.680	

Note: *ITP* = Initial task performances, *RTP* = Repeated task performances

Table 9 shows that for the Inter-cultural group there is a significant effect for repetition on the production of reactive LREs, $F(1,52) = 13.909$, $p = .0005$, $d = 1.097$, but no significant change for preemptive LREs. In contrast, intra-cultural repetition resulted in no significant change in the production of either LRE type.

Complexity

Table 10 shows the total LRE frequency counts categorized by complexity (simple and complex) generated by each pair in initial and repeated task performances.

Table 10. Descriptive statistics for LREs (complexity)

	Initial task performance			Repeated task performance		
LRE outcome	***M***	***SD***	***%***	***M***	***SD***	***%***
Inter-cultural group						
Simple	10.56	4.84	53.69	14.44	4.67	46.25
Complex	9.11	4.78	46.31	16.78	6.23	53.75
Total LREs	19.67	7.95	100	31.22	7.01	100
Intra-cultural group						
Simple	10.56	3.02	58.32	8.33	2.64	58.13
Complex	7.56	4.64	41.74	6.00	4.48	41.87
Total LREs	18.11	4.94	100	14.33	4.63	100

The above table reveals that for the Intra-cultural group simple LREs were more common than complex LREs during both the initial and repeated performances. Conversely, for the Inter-cultural group, complex LREs outnumbered simple LREs during the repeated task performances.

To investigate the effect of repetition on changes in the complexity of LREs for each group, a one-way MANOVA was conducted for each group separately. The results of the test of between-subjects effects are shown in Table 11.

Table 11. Test of between-subjects effects for repetition (complexity)

Variable	***F***	***df***	***Sig.***	**Direction**
Inter-cultural group				
Simple	4.041	1	.051	
Complex	10.851	1	.002	RTP > ITP
Intra-cultural group				
Simple	2.735	1	.107	
Complex	.636	1	.431	

Note: *ITP* = Initial task performance, *RTP* = Repeated task performance

Results from Table 11 indicate that for the Inter-cultural group there is a significant effect for repetition on complex LREs, $F(1,52) = 10.851$, $p = .002$, $d = .969$. Conversely, inter-cultural repetition resulted in no significant change in the production of simple or complex LREs.

Resolution

Table 12 shows the total LRE frequency counts for each category of resolution.

Table 12. Descriptive statistics for LREs (resolution)

	Initial task performance			Repeated task performance		
LRE resolution	***M***	***SD***	***%***	***M***	***SD***	***%***
Inter-cultural group						
With uptake	5.00	2.40	25.41	11	5.24	35.23
Without	12.56	5.66	63.85	18.44	5.69	58.94
Incorrect	1.33	1.50	6.76	.50	.69	1.60
Unresolved	.78	.89	3.97	1.28	1.18	4.10
Total LREs	18.11	4.94	100	14.33	4.63	100
Intra-cultural group						
With uptake	5.11	2.97	28.23	5.22	3.29	36.43
Without	10.77	3.70	59.12	7.33	2.16	51.15
Incorrect	1.44	1.59	7.96	1.22	1.33	8.51
Unresolved	.78	.88	4.31	.56	.80	3.91
Total LREs	19.67	7.95	100	31.22	7.01	100

The above table reveals that the majority of LREs (51–63%) were resolved without uptake. That is, most LREs reached a correct solution but with no evidence of uptake.

A one-way MANOVA was conducted to test for the effect of repetition. Table 13 presents the results of the test of between-subjects effects.

Table 13. Test of between-subjects effects for repetition (resolution)

Variable	***F***	***Df***	***Sig.***	**Direction**
Inter-cultural group				
With uptake	10.894	1	.002	RTP > ITP
Without	6.457	1	.014	RTP > ITP
Incorrect	3.788	1	.058	
Unresolved	1.275	1	.264	
Intra-cultural group				
With uptake	.006	1	.939	
Without	5.525	1	.025	ITP > RTP
Incorrect	.105	1	.748	
Unresolved	.318	1	.577	

Note: *ITP* = Initial task performance, *RTP* = Repeated task performance

Results from Table 13 indicate that for the Inter-cultural group there is a significant effect for repetition on resolved LREs with uptake, $F(1,52) = 10.894$, $p = .002$, $d = .971$,

and resolved LREs without uptake, $F(1,52) = 6.457$, $p = .014$, $d = .748$, but no significant change in the number of LREs that were categorized as incorrect or unresolved. That is, inter-cultural repetition has significantly increased the production of both kinds of resolved LREs. Although inter-cultural repetition did not significantly impact LREs that were resolved incorrectly or left unresolved, proportionally these weighed less heavily in the repeated performances. Finally, intra-cultural repetition resulted in a significant decrease in the number of LREs that were resolved without uptake, $F(1,52) = 5.525$, $p = .025$, $d = .806$.

Discussion

The results of this study shed light on two interrelated and under-researched areas of task repetition: the impact of task repetition and of interlocutor variation on learners' attention to linguistic form. Whereas Aubrey (2017) showed that learners are more emotionally engaged during inter-cultural task-based interactions, this follow-up study provides evidence that such interactions can also induce more frequent episodes of linguistic problem-solving.

Specifically, this study investigated the effect of task repetition on LRE production when a repeated performance is practiced in intra-cultural (RQ1) and inter-cultural pairs (RQ2). Intra-cultural repetition had a small (i.e. non-significant) negative effect on LRE production. In contrast, inter-cultural repetition had the effect of significantly increasing the number of LREs produced overall ($p = .041$, $d = 1.462$) and for each individual task within each attempt (Task 1/1R: $p = .031$, $d = .971$; Task 4/4R: $p = .016$, $d = .1.096$). Table 14 summarizes the effect of task repetition on linguistic focus, type, complexity, and resolution for LREs for each group.

It is interesting to note that task repetition did not increase the number of lexical LREs in any significant way. This suggests that, regardless of interlocutor, learners reduce their need to negotiate the meanings of words or phrases when repeating tasks. This can be explained in terms of changes in allocation of limited attentional resources to language during the repeated task performances (Skehan, 1998). During the initial task performances, learners dedicated their attention to resolving lexical issues as these had the highest value when conveying meaning. In fact, more than 50% of LREs had a lexical focus for both groups. During the repeated task performances, learners presumably benefited from the knowledge of lexical items acquired during the initial performances, which may have aided in both the clarity of message formulation and the ability to interpret intended meaning, diminishing the need for lexical LREs.

An important issue, therefore, is how learners utilized their 'freed up' attention during the repeated task performances. For the Intra-cultural group, it seems that learners did not turn their attention to linguistic form – neither by producing grammatical

Table 14. Summary of significance levels for dependent variables

	Inter-cultural group		Intra-cultural group	
	Significance	**Direction**	**Significance**	**Direction**
Linguistic focus				
Lexical	*Non-sig.*	–	.042	ITP > RTP
Grammatical	.003	RTP > ITP	*Non-sig.*	–
Sp./Pron.	.025	RTP > ITP	*Non-sig.*	–
Other	*Non-sig.*	–	*Non-sig.*	–
Type				
Reactive	>.001	RTP > ITP	*Non-sig.*	–
Preemptive	*Non-sig.*	–	*Non-sig.*	–
Complexity				
Simple	.050	ITP > RTP	*Non-sig.*	–
Complex	.002	ITP > RTP	Non-sig.	–
Resolution				
With uptake	.002	RTP > ITP	*Non-sig.*	–
Without uptake	.014	RTP > ITP	.025	ITP > Task P2
Incorrect	*Non-sig.*	–	*Non-sig.*	–
Unresolved	*Non-sig.*	–	*Non-sig.*	–

Note: *Sp./Pron.* = Spelling/Pronunciation, *ITP* = Initial task performances, *RTP* = Repeated task performances

LREs nor spelling/pronunciation LREs. The lack of grammatical problem-solving on the part of Japanese learners may be a result of insufficient L2 knowledge and/or confidence. In other words, they may have lacked the confidence needed to enter into negotiations about grammar and/or failed to notice errors made by their interlocutor. Example (6) shows how a grammatical LRE may have been either avoided or missed.

Example 6. (Intra-cultural group, Task 1R)

```
1  Goro:  My name is Goro.
2  Aoi:   Goro, would you like…visit?
3  Goro:  London in British.
4  Aoi:   I want to visit Copenhagen in Denmark.
```

Example (6) illustrates how two Japanese learners began their task. Aoi struggles to ask the question "where would you like to visit?", and instead asks, "Goro, would you like… visit?" (turn 2). Although no data were collected on the thought processes of these learners, one could surmise that Goro understood this utterance because he possesses a high degree of empathy as (1) a fellow Japanese learner and (2) a student who has completed this task previously, enabling him to make an accurate prediction

of the communicative intention of his interlocutor. Alternatively, cultural motivations, such as avoiding face-threatening linguistic behavior – a characteristic of Japanese communication (see Fuji & Mackey, 2009) – could explain why some intra-cultural interactions resulted in the avoidance of LREs. In sum, though task repetition may have made it easier for learners to express their intended meaning, it did not direct their attention to linguistic form.

Quite the opposite is true for the Inter-cultural group. For the repeated task performances, instead of increasing their attention to meaning-focused (i.e. lexical) episodes, the inter-cultural pairs devoted more attention to form, significantly increasing the number of grammatical and spelling/pronunciation LREs. The international students clearly demonstrated that they were capable of noticing and willing to provide feedback on grammatical errors. Example (7) shows how an international student provided grammatical feedback.

Example 7. (Inter-cultural group, Task 2R)

66 Katsu: And everyone play football.
67 Malori: Everybody CAN- CAN play football?
68 Katsu: Yeah, so you… what do you think?

Unlike Example (6), here the Japanese learner's interlocutor (Malori) immediately provided a recast upon noticing the grammatical error. This kind of reactive LRE, in which the international interlocutor recasts a problematic utterance, was frequently observed in inter-cultural task performances. As noted previously, inter-cultural task repetition resulted in a significant increase in reactive LREs, with no significant change in preemptive LREs. In contrast to preemptive LREs, recasting a linguistic error requires both noticing there is a problem with another's production, and then drawing on one's linguistic resources to provide the correction. The abundance of reactive LREs is most likely due to the proficiency differences between pairs, but may also be attributed to the perceived role of the international student as the 'expert' English user.

Related to this is the impact of task repetition on the complexity of LREs. Inter-cultural repetition had the effect of significantly increasing the number of complex LREs ($p = .002$, $d = .969$). This large effect suggests that students spent more turns resolving each issue during repeated task performances. This has considerable implications for potential language learning. As Loewen (2005) claims, "the complexity or length of negotiation surrounding a linguistic item might influence the saliency of that item, with longer negotiation sequences being potentially more salient" (p. 366). Conversely, intra-cultural task repetition did not significantly change the number of complex LREs produced. This reflects Williams' (2001) claim that learner-learner interaction results in LREs that are relatively short as learners lack the linguistic resources needed to make their language-related discussions more complex.

A final interesting difference is the way in which task repetition affected the resolution of LREs. The Inter-cultural group significantly increased the amount of LREs that were resolved without uptake. This is not surprising considering the sheer amount of feedback that the international students provided. Furthermore, the Inter-cultural group also significantly increased the amount of LREs that were resolved with uptake. Resolving LREs with a demonstration of uptake requires both students to use the corrected form of the problematic language item. In the case that an international student provides feedback, the learner would need to incorporate the target structure into his/her subsequent utterance. The increase in the number of LREs in which uptake is demonstrated was perhaps aided by the fact that the number of reactive LREs also increased. As opposed to preemptive LREs, reactive LREs are triggered largely in reaction to a perceived error in an utterance, thus the resolution of reactive episodes serves the purpose of filling an observed linguistic gap, creating a context in which uptake is a more useful – and more likely – outcome for learners. Moreover, the greater complexity of LREs as a result of inter-cultural repetition means that learners have more turns to demonstrate uptake.

Intra-cultural repetition, on the other hand, had the effect of significantly lowering the number of LREs resolved without uptake and marginally (non-significantly) raising the number of LREs with uptake. In fact, the distribution of the number of LREs produced during the repeated task performances indicates that the proportion of LREs resolved with uptake increased from 28% to 36%, whereas the distribution of unresolved LREs remained constant (*ITP* = 12%; *RTP* = 12%). Thus, the drop in the number of LREs had the effect of slightly distorting the distribution of LRE resolution in a way that learners were more likely to demonstrate uptake when they had the chance. This result supports claims made by Mackey et al. (2007) who found that during task repetition, learners were able to use their feedback in subsequent utterances.

The comparatively high number of LREs generated by pairs in the inter-cultural condition may have been due to changes in task engagement. To support this view, it is useful to draw on the findings of Aubrey (2017) for insights into effects of inter-cultural contact and Kim (2013) for the effects of task repetition. Aubrey (2017) found that learners participating in intra-cultural interaction reported higher flow levels (*interest*, *challenge*, *control*, and *attention* during a task) and elevated levels of interactivity (significantly more turns of talk), which was argued to be partly driven by learners' curiosity to understand a new – and novel – cultural perspective (see Aubrey, 2017 for flow-inducing examples of interaction). The current study suggests that inter-cultural interaction also leads to heightened cognitive engagement around language issues. In regard to the effect of task repetition, Kim (2013) reported that learners who performed oral, collaborative tasks under the condition of exact task repetition produced significantly fewer LREs than those who performed tasks under the procedural repetition condition. She concluded that exact task repetition is not as effective as

procedural repetition in promoting linguistic discussions. To explain this, Kim drew a connection between the lack of attention to form and a decreased sense of "interest" and "likeability" of repeating the same tasks with the same procedures, and therefore pointed to the 'novelty' of performing a task with new content to explain the increase in students' willingness to enter into LREs. Perhaps a similar conclusion can be drawn here. The positive effect of performing a task with an international interlocutor and may have neutralized the negative effect of repeating a task with the same content and procedures. More research into the relationship between inter-cultural contact, task repetition and the different facets of task engagement is needed to support this claim.

Some limitations of this research need to be highlighted. First, this study did not attempt to tease out the relative impact of the interlocutors' proficiency levels and cultural backgrounds. Perhaps future studies can make use of a larger sample size to include proficiency as an independent variable. Second, the current study relied only on the use of transcripts to identify LREs and interpret their type. A more reliable picture of language negotiation would involve stimulated recall methods (as suggest by Swain & Lapkin, 1998), which may reveal participants' motivations to enter into such episodes. In addition, triangulating data with instruments designed to measure learners' engagement during tasks may shed of light on the emotional and motivational states surrounding identified language learning opportunities.

Conclusion

The aim of this study was to examine the effect of task repetition on students' attention to linguistic form under the conditions of intra-cultural and inter-cultural interaction. It has shown that when oral tasks are repeated with different interlocutors, language learning opportunities can change significantly. The findings suggest that the repetition of tasks with an interlocutor from a different cultural and linguistic background may have the effect of promoting language-related negotiations. In fact, the results of this study indicated that such repeated inter-cultural interactions were effective at generating complex and grammatically focused episodes, which tended to be resolved more often than during the initial task performances. In contrast, this study also shows that repeating a task with an interlocutor of a similar cultural and linguistic background does not result in a significant change in language negotiations. This research benefited from the methodological decision to expand on previous LRE classification frameworks, providing a more complex picture of intra-cultural and inter-cultural interaction.

From a pedagogical perspective, this research provides some evidence to support the practice of face-to-face inter-cultural task interaction in the EFL classroom. As this study was conducted at a university that is, in terms of proportion of international

students, not unlike many university contexts in Japan, it could also be considered an instructional template for how EFL teachers can utilize a relatively small pool of available international students for English-language learning purposes.

References

Ahmadian, M., & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity and fluency of EFL learners' oral production. *Language Teaching Research*, 15(1), 35–59. https://doi.org/10.1177/1362168810383329

Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman and Hall.

Aubrey, S. (2017). Inter-cultural contact and flow in a task-based Japanese EFL classroom. *Language Teaching Research*, 21(6), 717–734.

Bowles, M., Adams, R., & Toth, P. D. (2014). A comparison of L2-L2 and L2-heritage learner interactions in Spanish language classrooms. *Modern Language Journal*, 98(2), 497–517. https://doi.org/10.1111/modl.12086

Bygate, M. (1996). Effects of task repetition: Appraising the developing language of learners. In J. Willis & D. Willis (eds). *Challenge and change in language teaching* (pp. 136–146). Oxford: Heinmeann.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (eds.), *Researching pedagogic tasks, second language learning, teaching and testing* (pp. 23–48). Harlow: Longman.

Ellis, R., Basturkmen, S., & Loewen. (2001). Learner uptake in communicative ESL lessons. *Language Learning*, 51(2), 281–318. https://doi.org/10.1111/1467-9922.00156

Fernández Dobao, A. (2012). Collaborative dialogue in learner-learner and learner-native speaker interaction. *Applied Linguistics*, 33(3), 229–256. https://doi.org/10.1093/applin/ams002

Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York, NY: Wiley.

Fortune, A., & Thorp, D. (2001). Knotted and entangled: New light on the identification, classification and value of language related episodes in collaborative output tasks. *Language Awareness*, 10(2–3), 143–60. https://doi.org/10.1080/09658410108667031

Fuji, A., & Mackey, A. (2009). Interactional feedback in learner-learner interactions in a task-based EFL classroom. *International Review of Applied Linguistics in Language Teaching*, 47(3–4), 267–301

Kim, Y. (2013). Promoting attention to form through task repetition in a Korean EFL context. In K. McDonough & A. Mackey (Eds.), *Second language interaction in diverse educational settings* (pp. 3–24). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.34.04ch1

Kim, Y., & Tracy-Ventura, N. (2013). The role of task repetition in L2 performance development: What needs to be repeated during task-based interaction? *System*, 41(3), 829–840. https://doi.org/10.1016/j.system.2013.08.005

Kim, Y., & McDonough, K. (2008). The effect of interlocutor proficiency on the collaborative dialogue between Korean as a second language learners. *Language Teaching Research*, 12(2), 211–234. https://doi.org/10.1177/1362168807086288

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. https://doi.org/10.2307/2529310

Leeser, M. (2004). Learner proficiency and focus on form during collaborative dialogue. *Language Teaching Research*, 8(1), 55–81. https://doi.org/10.1191/1362168804lr1340a

Loewen, S. (2005). Incidental focus on form and second language learning. *Studies in Second Language Acquisition*, 27(3), 361–86. https://doi.org/10.1017/S0272263105050163

Long, M. (1991). Focus on form: a design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39–52). Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.2.07lon

Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, 4(3), 221–250.

Mackey, A., Kanganas, A. P., & Oliver, R. (2007). Task familiarity and interactional feedback in child ESL classrooms. *TESOL Quarterly*, 41(2), 285–312. https://doi.org/10.1002/j.1545-7249.2007.tb00060.x

Manchon, R. (2014). The distinctive nature of task repetition in writing: Implications for theory, research and pedagogy. *Esudios de Linguistica Inglesa Aplicada*, 14, 13–41.

Patanasorn, C. (2010). *Effects of procedural content and task repetition on accuracy and fluency in an EFL context* (Unpublished PhD dissertation). Northern Arizona University.

Skehan, P. 1998. *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P., & Foster, P. 1999. The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49(1), 93–120. https://doi.org/10.1111/1467-9922.00071

Storch, N., & Aldosari, A. (2013). Pairing learners in pair work activity. *Language Teaching Research*, 17(1), 31–48. https://doi.org/10.1177/1362168812457530

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–256). Cambridge, MA: Newbury House.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honor of H. G. Widdowson* (pp. 125–144). Oxford: Oxford University Press.

Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 94–119). Oxford: Oxford University Press.

Swain, M. (2005). Legislation by hypothesis: The case of task-based instruction. *Applied Linguistics*, 26(3), 376–401. https://doi.org/10.1093/applin/ami013

Swain, M. (2006). Languaging, agency and collaboration in advanced second language proficiency. In H. Bynes (Ed.), *Advanced language learning: The contribution of Halliday and Vygotsky* (pp. 95–108). London: Continuum.

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *Modern Language Journal*, 82(3), 320–337. https://doi.org/10.1111/j.1540-4781.1998.tb01209.x

Watanabe, Y., & Swain, M. (2007). Effects of proficiency differences and patterns of pair interaction on second language learning: Collaborative dialogue between adult ESL learners. *Language Teaching Research*, 11(2), 121–142. https://doi.org/10.1177/1362168806074599

Williams, J. (2001). The effectiveness of spontaneous attention to form. *System*, 29(3), 325–340. https://doi.org/10.1016/S0346-251X(01)00022-7

Yule, G., & Macdonald, D. (1990). Resolving referential conflicts in L2 interaction: The effect of proficiency and interaction role. *Language Learning*, 40(4), 539–536. https://doi.org/10.1111/j.1467-1770.1990.tb00605.x

CHAPTER 6

Effects of task type, task-type repetition, and performance criteria on L2 oral production

Xingchao Hu
Nanjing Medical University, China

This study investigates the impact of task type, task-type repetition, and performance criteria on oral second language production. One hundred and forty-four Chinese learners of English were randomly paired and assigned to either a map task or a picture-story task. Each task type was repeated a week later via a different exemplar. Each experimental group was then further divided between "high" and "low" performance criterion groups, and a control group (CG). Both the high criterion (HC) and the low criterion (LC) groups were informed at the outset that they would be asked to perform an additional activity upon completion of the baseline task. The HC speakers undertook a more demanding post-task activity and the LC group a less demanding one, with no extra post-task requirement for the CG. Analysing participants' language production in terms of complexity, accuracy, and fluency, a strong effect for task type was observed. Task-type repetition resulted in greater accuracy and affected all fluency measures (though not all in the same direction). Finally, the high criterion condition was positively related to phrasal complexity and fluency on both task types, but negatively associated with subordination complexity on the picture-story task, findings not attested in previous studies. Results are discussed in terms of the limited attentional resources model on second language performance. The study contributes to our understanding of how pre-specified task performance criteria can affect students' performance of a task.

Introduction

Although there are different approaches to task-based research, this study adopts a cognitive approach. This approach focuses in particular on three aspects of language production: complexity, accuracy, and fluency. Drawing on a limited processing capacity theory, it proposes that language users differ in the extent to which they prioritise the three aspects of language production, which can lead to performance on one aspect improving at the expense of others, what are termed trade-off effects (Foster & Skehan, 1996). The trade-off effect can respond to factors such as task features,

DOI 10.1075/tblt.11.06hu

implementation conditions, and individual learning styles. Although the limited capacity position has been challenged by Robinson (2003, 2005), arguing that it is possible for learners to simultaneously attend to more than one aspect of language and language processing, as Ellis (2009) has noted, the results of previous task-based research are overall more supportive of Skehan's cognitive position. It is this perspective of second language use and processing that the current study adopts.

Background

Task type

A wide range of task types has been employed in task research, where the term *task type* is generally closely related to discourse features or functional demands required by a particular task, such as giving instructions/directions (e.g., Crookes, 1989; Foster & Skehan, 1996; Mehnert, 1998; Sangarun, 2005), making decisions (Foster & Skehan, 1996; Skehan & Foster, 2005), interviews (Bygate, 2001; Tarone, 1985), general discussion (Wigglesworth, 1997), exposition involving making apologies and giving excuses (Mehnert, 1998), academic lecturettes (Lynch & Maclean, 2000, 2001), and argumentations (Bygate, 1999; Sangarun, 2005). The most frequently employed task is narratives (story-telling usually elicited either by pictures or films) (e.g., Bygate, 1996, 2001; Ellis & Yuan, 2005; Gass et al., 1999; Ortega, 1999; Robinson, 1995; Skehan & Foster, 1999; Tavakoli & Foster, 2008; Tavakoli & Skehan, 2005; Yuan & Ellis, 2003).

Numerous studies have compared the quality of oral production on different types of tasks. There is evidence that variability in task design is significantly related to different aspects of language performance. For example, task designs which require more attention to language form, impose greater communication pressure, or demand higher textual cohesiveness have tended to result in greater accuracy in certain grammatical forms (Tarone, 1985; Tarone & Parrish, 1988). By contrast, tasks with a clear structure involving transmitting personal and familiar information have been more likely to promote fluency and accuracy at the cost of complexity (Foster & Skehan, 1996; Mehnert, 1998). Bygate (1999) compared the grammatical patterns of learner language on narratives and argumentations and reported that the narratives gave rise to more verb arguments and more frequent use of relative clauses, whereas the argumentations resulted in more verb groups, individual verb forms, and nominal clauses. As he noted, this is because the narratives mainly concerned identifying referents and following events and time relations, whereas the argumentations concentrated on presenting reasons and justifications. Another study by Bygate (2001) examined how narratives and interviews might potentially influence language production. The results showed that the interview tasks resulted in significantly less fluency and greater

complexity than the narrative tasks, which, Bygate (2001) suggested, was related to the interpersonal aspect of the interview tasks, where the interlocutor might cause the speaker "more unexpected on-line processing" and encourage more attention on the part of speakers to develop greater complexity (p. 38). From a slightly different perspective, Derwing et al. (2004) examined whether raters' assessment of fluency varied across tasks. They found that the degree of freedom speakers had in choosing lexis, syntax and content while performing a given task may also influence their language output, with greater freedom leading to greater perceived fluency. Although the influence of task-type difference on linguistic production has been explored in a number of previous studies, Ellis (2009) in a recent review article pointed out that this is still an under-researched area.

Task rehearsal

In addition to the various conditions mentioned above that can be manipulated before and while a task is performed, task rehearsal (referred to as "task iteration" by Larsen-Freeman, this volume) has also been investigated. This includes task repetition (repeating an identical task), task-type repetition (repeating a similar task of the same type), and task recycling (repeating a specific task with different interlocutors).

Previous studies have shown some evidence that task repetition can impact on L2 oral production, whereas the effect of task-type repetition has been less clear (but see Kim et al. this volume). For example, Bygate (1996) asked one learner to perform a narrative task on two separate occasions and found that overall the learner's second performance was better than the first one in terms of accuracy, lexical and grammatical repertoire, and fluency. Bygate attributed these findings to the fact that, whereas on the first occasion more of the speaker's focus of attention was on conceptualising the informational content, on the second occasion attention could be shifted away from conceptualizing towards encoding the message, monitoring the language and organising the discourse. In a later study Bygate (2001) examined how task repetition and task-type repetition might affect language use on two types of tasks – narratives and interviews. The results showed that task repetition was significantly related to increased complexity on both types of task, whereas fluency was negatively affected on the repeated interview task. Once again Bygate (2001) reasoned that "the highly contextualised cognitive rehearsal" helped make spare attentional resources available to focus on the formulation of the message content (p. 42). In terms of task-type repetition, only a partial practice effect was observed. In other words, language production improved on specific practiced tasks to which speakers had been exposed over a period of time rather than on similar types of tasks. This might have to do with "some kind of build-up from the experience of attending to tasks of the same type" (Bygate, 2001, p. 42).

Gass et al. (1999) investigated learners' linguistic output on repeating tasks of the same content (task repetition) and repeating tasks of different content (task-type repetition). Learners in their study exhibited better performance in terms of overall proficiency on repeated tasks. Speakers repeating the same-content task, in particular, demonstrated a greater increase in accuracy and lexical sophistication. These findings suggested that repeating a similar task of the same type, particularly repeating an identical task was, in some way, beneficial to language production.

Lynch and Maclean (2000, 2001) were interested in the influence of task recycling (same task with different interlocutors) on learners' oral performance. They found that the speakers, over six cycles, were able to produce more accurate language in terms of phonology, lexical access or selection, semantic precision, and syntax, although the type of improvement in their performance depended on the learners and their proficiency levels.

On the whole previous studies have assumed that task performance is largely shaped by task design, by interlocutor, and by performance conditions such as availability of planning time. However another possible factor influencing the way learners carry out tasks is the standard of performance that the learners think is expected of them – in other words, how precisely are they trying to do it? This study investigates this possibility by exploring whether learners' performance on a given task is affected by the goals set for them on previous encouters with a similar task. In other words, does the level of demand contained in a pre-task briefing affect performance not only on the immediate task, but on subsquent similar tasks?

Task performance criteria

Unlike followers of the cognitive approach, researchers who take a sociocultural position on task performance argue that language production or use varies depending on the interaction between individuals and the tasks they engage in, rather than on the inherent properties of the task itself (Appel & Lantolf, 1994). Coughlan and Duff (1994), for instance, reported that the same task can result in different activities when performed by different learners or even by the same learner on different occasions. This is because when engaging in a task, learners bring their own purposes and establish their own goals so as to orient themselves to how the task can be completed, and in the process have to reinterpret and redefine the task, which will potentially influence their language. From the perspective of collaborative theory, successful communication is considered to depend on a joint process of achieving *grounding*, namely, the shared assumption which speakers work on that what is said is understood (Clark & Brennan, 1991; Clark & Schaefer, 1989). In the process of achieving grounding, people do not stop and seek perfect understanding all the time, but rather an understanding "to a criterion sufficient for current purposes" (Clark & Wilkes-Gibbs, 1986). This is called the

criterion of grounding. Different goals of communication, types of conversation, and other factors, such as the medium of communication all result in different criteria of grounding. For example, Clark and Brennan (1991) noted that task-oriented conversation would likely have a higher criterion of grounding than free casual conversation. Essentially, grounding criteria can be viewed as a necessary basis for task performance.

Although collaborative theory is not directly related to the present study, the concept of grounding criteria, or in other words, task performance criteria, can shed light on how L2 learners perform a given task. For instance, if a task is implemented with no particular performance criterion, speakers are bound to set their own, and these will inevitably vary across different task types and participants. This is consistent with claims made from the sociocultural perspective that while engaging in a particular task, participants' individual expectations and intentions, purposes, and goals in the task may all shape the way they conceptualise and conduct the task, leading to different performances. Furthermore, learners' performance on a given task will vary when different performance criteria are specified by the teacher or task designer. Tannen's (1980) study suggested how task performance was affected by performance criteria implicitly set by participants from different cultural backgrounds. She compared the ways in which Greek and North American speakers performed a narrative task and found that the Greek speakers treated the retelling task as a kind of literary or dramatic challenge and were concerned about how to make the retelling more interesting. By contrast, the North American speakers spontaneously interpreted and treated the task as a kind of psychology or memory test and concentrated on achieving accurate recall.

Wilkes-Gibbs (1997), on the other hand, described how communication can be affected by different performance criteria externally specified by the researcher. Participants in her study were asked to perform a map task by describing the entire tour route either in order to enable another person to go through it from memory (high criterion) or simply to enable them to estimate travel time (low criterion). The results showed that the high criterion dyads conversed more extensively about the tour route and spent significantly more time on the task than their low criterion peers. This study also investigated the participants' actual understanding by asking them to draw the complete tour route on a new map at the end of the session. Surprisingly, the high criterion pairs did not outperform their low criterion counterparts. One plausible account provided by the researcher was that the final task of drawing a new map from memory was either too easy or too demanding, that is, not sufficiently discriminating in nature to distinguish performances between the high criterion and low criterion dyads.

Wilkes-Gibbs' study is not concerned with pedagogical intervention. However, if outcome criteria can impact on task performance in a research project described above, the question arises whether the same effect obtains when tasks are used for pedagogical purposes, and this is a third factor explored in the present study.

The study

Most examples of tasks in the research and pedagogical literature are presented and discussed as if performance criteria were irrelevant. The present study, therefore, intended to address that gap by investigating the impact of different levels of task criteria on L2 speakers' language production. Performance criteria were operationalised by implementing the same baseline task but with post-task activities that differed in the outcome they demanded of the participants. The purpose was to lead learners to accept a particular criterion level, reflecting the specific goal that they were set. It could be argued that introducing externally manipulated performance criteria by giving students different task instructions would actually lead learners to perform different tasks, since they are bound to perceive the baseline task in different ways and engage in different cognitive processes. However, it was presumed in this study that the baseline task was neutral, but that students would inevitably import their own performance criteria, and so externally provided performance criteria were seen as task design factors which could be manipulated along a scale (with different levels). Following this line of thinking, for the purposes of either research or pedagogy, pressure on learners can be relaxed by underspecifying a task in terms of performance criteria, to the point that the criteria may be left open to the learners' own interpretation (as is often the case both in research and in course books). On the other hand, pressure on speakers can also be raised by higher performance criteria.

A second focus of this study was the effect of repeating a similar task of the same type. The approach here was slightly different from task-type repetition examined in other studies (e.g., Bygate, 2001; Gass et al., 1999). The intention was not only to run a different version of the same task, but to apply similar performance criterion conditions as on the first encounter. Hence, both task and performance criteria were repeated.

Finally, as Ellis (2009) noted, the effect of task type on language performance is still an under-researched area (Ellis, 2009). It was therefore useful to employ two types of tasks in this study. Performance criteria were studied within the context of the repetition of two task types, enabling us to explore the impact of performance criteria interacting with task type and task repetition.

Independent variables

Task type

Two sets of tasks (maps and picture-stories) were used for the experimental study. Each task set consisted of a warm-up task and two experimental tasks. In the case of the map task, the participants were asked to complete a tour route and locate all the

target places on their map by exchanging information with each other. In the case of the picture-story task, each member of the dyad was allocated three different pictures. The pairs were expected to tell each other what the pictures were about and then put them in the right order to make up a coherent story.

In the meantime, it was crucial that the different versions within the same task set were equivalent so that any significant differences in language performance were not attributable to the differences between the two versions of the same type of task. To ensure comparability, a series of independent samples t-tests were conducted separately (for the two sets of tasks) on language production with the control group on their first encounter with each task version. No significant differences were yielded on any of the measures for complexity, accuracy, and fluency, indicating that the two exemplars employed for both types of tasks were equivalent in terms of complexity level.

There were two main reasons why these tasks were chosen. Firstly, these types of tasks were widely used in task-based research, thus making it possible to compare findings with those obtained from other studies. More importantly, given that the two types of tasks differed in terms of communicational goals and functional requirements, they were expected to yield differences in language, notably in terms of discourse features. The map task mainly involved following a route to locate referents by giving and receiving directions, where speakers were likely to utter a large number of short and fairly simple imperative sentences. With limited communicational goals, the map task might also create a context where learners found it sufficient for their purposes to use repeatedly certain vocabulary items and syntactic patterns (e.g., direction-giving and location identification). In contrast, the narrative picture-story task required speakers to describe and sequence a series of pictures, which was "more concerned with identification of referents (who did what to whom)", where "the use of relative clauses" and "the use of adverbials to specify location and sequential occurrence" were more likely to occur (Bygate, 1999, p. 202). In addition, complex noun phrases and a large variety of attributive adjectives were also important for distinguishing referents in the picture-story task (Brown & Yule, 1983), but were less likely to be used in the map task. Further, in the picture-story task, speakers needed to use reasoning skills to relate the pictures they could see to those held by their interlocutors so as to work out a logical story line. When their reasons for interpreting why one event followed the other were challenged or rejected, a justification needed to be provided. To satisfy these reasoning demands, speakers were more likely to employ specific linguistic features, including "use of logical connectors (*if-then, so, therefore, because, etc.*)" (Robinson, 2001, p. 31) and "use of psychological, cognitive state verbs (*e.g., know, believe, suppose, think*)" (Robinson, 2005, p. 5). Hence, it could be anticipated that the picture-story task might lead to greater complexity compared with the map task.

Task-type repetition

The participants under different implementation conditions were requested to perform two versions of the same type of task at separate times with an interval of one week. They were not informed beforehand that they would be asked to perform a similar task on a second occasion. In the meantime, the participants did not receive any additional L2 input from the researcher apart from their usual language classes. It should be noted that during the second encounter, both task type and implementation conditions were held constant. In other words, since an element of performance criterion was built into the design, the two experimental groups did not simply perform a similar baseline task on the second occasion, but undertook it with a different pre-set (high vs. low) performance criterion.

Drawing on Logan's (1988) instance theory, it can be argued that after the first experience with a particular task, information concerning the nature and requirement of that task, as well as the steps and procedures to complete it, is encoded and stored in participants' long-term memory. When a same or similar task is repeated, the corresponding schemata is activated and retrieved, enabling more rapid and efficient resolution of the current task. Similarly, from a discourse analysis perspective, Bygate (2001, p. 29) pointed out that learners' previous experience of dealing with a discourse type contributes to their communicative competence and effectiveness in coping with a new exemplar of the same type. He further suggested that different versions of one discourse type share certain features, and that practicing a familiar and a novel task may affect the amount of attentional resources allocated to different aspects of speech processing, resulting in differentiated performances on the two tasks. It was therefore hypothesised that repeating a similar task of the same type might lead to improved accuracy and/or fluency.

Task performance criteria

Three levels of task performance criteria were established: (1) high criterion (HC), (2) low criterion (LC), and (3) zero criterion (control group), with differentiated levels of post-task requirements imposed on the two experimental groups and no additional demand assigned to the control group (see Table 1). In terms of the map task, the control group (CG) was simply asked to perform the baseline task of completing the travel route which connected a number of places to visit on their map. The HC participants, however, were told in advance that upon completing the baseline task, they would be asked to reproduce individually the whole map by marking the entire tour route together with all the places (both visited and not visited) on a blank map. The LC participants were informed at the outset that after finishing the primary task, they would be required to report individually to a third person all the places they had

visited. With regard to the picture-story task, the CG was simply asked to put their six pictures in the correct order based on the story line they had worked out. For the two experimental groups, upon completion of the baseline task, the HC speakers were asked to draw the other three pictures held by their interlocutors, whereas the LC participants were required to tell the whole story to a third person individually.

Table 1. Description of task performance conditions

	Conditions		
	High criterion (HC)	**Low criterion (LC)**	**Control group (CG)**
Map task	Reproducing the entire map by marking the tour route and all the places	Reporting the visited places to a third person	No post-task requirement
Picture-story task	Drawing the three pictures held by their interlocutors	Retelling the whole story to a third person	No post-task requirement

As we can see, in the experimental groups, successful completion of the post-task activity required both the HC and LC speakers to memorise and transfer information they themselves held, as well as that obtained from their interlocutors during the performance of the baseline task. However, the amount of information and the degree of precision demanded by the two post-task activities were intended to differ, the assumption being that the higher the criterion, the greater amount of and the more precise the information the speakers would need. Therefore, it was anticipated that the CG who performed the baseline task with no extra post-task requirement would retain the strategy of *skeletonizing* in their communication, where "only the bare events of a narrative are provided". The HC task performers, on the other hand, would be more likely to employ a highly *embroidered* strategy in which "these same events are described with a greater amount of expansion and embellishment" (Ehrlich et al., 1989, p. 403). However, little discussion has been made on the impact of different performance criteria on speakers' language use and processing. This was explored in the present investigation.

Research questions

The current study aimed to address the following research questions:

RQ 1: Is there a significant relationship between task type and language production?
RQ 2. Is there a significant relationship between task-type repetition and language production?

RQ 3: Is there a significant relationship between performance criteria and language production?

Method

Design

The study adopted a (2 × 2 × 3) mixed-method design (see Table 2). *Task* (2 types of tasks) and *condition* (3 levels of task performance criterion) were the between-subject variables, and *time* (2 encounters with similar tasks of the same type) was the within-subject variable.

Table 2. Design

	Conditions		
	High criterion (HC)	**Low criterion (LC)**	**Control group (CG)**
Map task	Time 1	Time 1	Time 1
	Time 2	Time 2	Time 2
Picture-story task	Time 1	Time 1	Time 1
	Time 2	Time 2	Time 2

Dependent variables

Complexity measures

In SLA research, a frequently used metric for syntactic complexity is to measure the amount of subordination. However, Norris and Ortega (2009) pointed out that complexity in itself is a complex subsystem with multiple dimensions. Subordination metrics alone might lead to misinterpretations of performance, especially by learners with advanced levels of proficiency, because a decrease in subordination might be connected with an increase in phrasal or overall complexity in their language production. As a result, Norris and Ortega (2009) argued that "syntactic complexity must be measured multidimensionally" (p. 462). Following their proposal, three dimensions of syntactic complexity were examined separately in the present study, namely, mean length of AS-unit for global complexity, clauses per AS-unit for complexity by subordination, and mean length of clause for phrasal complexity.

In terms of lexical diversity, the most well-known measure is probably type-token ratio (TTR), which is notorious for being highly sensitive to the length of texts. To address this shortcoming, an alternative measure, mean segmental type-token ratio

(MSTTR), was used to evaluate lexical richness. The transcript of each dyad's language output was divided into segments of 40 words and the MSTTR for each dyad was obtained by adding the TTR scores for their chunks and dividing the total by the total number of segments (see Yuan & Ellis, 2003).

Accuracy measure

Accuracy was measured by calculating the number of error-free clauses as a percentage of the total number of clauses.

Fluency measures

Fluency was also gauged multi-dimensionally, with percentage of mid-clause pauses as an index for breakdown fluency, speech rate and mean length of run for speed fluency, and percentage of dysfluencies as a measure of repair fluency (see Skehan & Foster, 2005).

Participants

The participants were 144 first-year undergraduate students from a Chinese university. At the time the data were collected, they were all between the ages of 18 and 19, with a history of six years of learning English in middle schools. Their scores in the National College Entry Examination for English were between 110 and 130 (a full score being 150). This was roughly equivalent to an intermediate level of English. None of the participants had ever visited an English-speaking country. They had had few opportunities to communicate in English outside the classroom. The participants could be regarded as a fairly homogeneous group in terms of their proficiency level and learning experience.

Data collection and coding

The 144 participants were randomly paired and divided into two task groups, performing either the map task or the picture-story task. In each task group, the students were further allocated at random to one of the three conditions, high criterion (HC), low criterion (LC), or control group (CG). Three different levels of task performance criterion were correspondingly set prior to the implementation of the tasks. Participants met twice for data collection. During the first session, each dyad was assigned the first experimental task immediately after completing a task-matched warm-up task (either

a map or picture-story task with different content from those used in the study). In the second session, one week later, another experimental task was implemented, which again was of the same type as the first, but different in content.

The conversations of each dyad during their performance of two similar tasks were recorded and fully transcribed as Word documents. Decisions were made to disregard chuckles, coughs, and other noises during the process of transcription. Utterances produced in the speakers' first language were translated into English, but were not included in the analysis. Filled pauses mostly transcribed as *uh* or *um* were counted as speech.

All speech data were coded for AS-units (Fosters et al., 2000), subordinate clauses, repetitions, self-corrections, and error-free clauses. A silence of 0.3 seconds or greater was counted as an instance of hesitation. The examination of complexity was carried out based on the pruned transcripts. This was because inclusion of certain data could distort the characteristics of language or the nature of performance. In order to ensure inter-rater reliability, a second rater was invited to code a randomly selected 10% of the speech data. The result of inter-rater reliability calculated by means of percentage agreement reached 94%.

Statistical analyses

A series of three-way mixed within-between subjects analysis of variance (ANOVA) procedures were performed separately for each dependent variable. The ANOVA analyses were then followed by Bonferroni post-hoc multiple comparisons when statistical significance was reached. If a significant interaction was observed between any two or the three independent variables, the three-way mixed ANOVA was further broken down into a series of one-way or two-way ANOVAs in order to locate the source of significance.

Results

This section reports the results for each of the three research questions.

Task type

Task-type difference had a significant effect on all complexity measures in favour of the picture-story task (see Table 3). With respect to global complexity, the picture-story task generated a strikingly higher mean length of AS-unit than the map task

with a difference of approximately 2.5 words per AS-unit ($F = 138.699$, $p < .001$). This pattern was repeated for both subordination and phrasal complexity, where the picture-story task produced significantly more clauses per AS-unit ($F = 94.818$, $p < .001$) as well as significantly longer clauses ($F = 78.901$, $p < .001$) compared with the map task. The picture-story task was also found to give rise to greater lexical diversity with a higher MSTTR ($F = 68.313$, $p < .001$).

Table 3. Effects of task type on complexity

	Time 1		Time 2	
	Mean	**SD**	**Mean**	**SD**
Global complexity (mean length of AS-unit)				
Picture-story	7.30	1.44	7.21	1.06
Map	4.85	0.77	4.85	0.62
$F = 138.699$, $p < .001$*				
Complexity by subordination (clause/AS-unit ratio)				
Picture-story	1.18	0.18	1.18	0.15
Map	0.90	0.13	0.91	0.12
$F = 94.818$, $p < .001$*				
Phrasal complexity (mean length of clause)				
Picture-story	5.85	0.79	5.84	0.69
Map	4.82	0.53	4.76	0.41
$F = 78.901$, $p < .001$*				
Lexical diversity (mean segmental type-token ratio)				
Picture-story	0.66	0.04	0.65	0.04
Map	0.60	0.03	0.59	0.04
$F = 68.313$, $p < .001$*				

Table 4 indicates that task type was also associated with accuracy where map task performers tended to generate at least 7% more accurate language than their picture-story counterparts ($F = 15.860$, $p < .001$).

In addition, task type was significantly related to fluency, favouring the map task in terms of mid-clause pauses and dysfluencies (Table 5). More specifically, the map task performers were more fluent than their picture-story peers by producing fewer interclausal pauses ($F = 244.319$, $p < .001$) and repeating and correcting themselves less frequently ($F = 25.936$, $p < .001$).

Table 4. Effects of task type on accuracy

	Time 1		Time 2	
	Mean	**SD**	**Mean**	**SD**
Error-free clauses (%)				
Picture-story	53.93	10.32	58.20	8.33
Map	63.08	9.14	65.56	10.55
	$F = 15.860$, $p < .001$*			

Table 5. Effects of task type on fluency

	Time 1		Time 2	
	Mean	**SD**	**Mean**	**SD**
Breakdown fluency (mid-clause pauses %)				
Picture-story	55.97	8.61	52.86	8.55
Map	35.36	6.17	31.82	4.76
	$F = 244.319$, $p < .001$*			
Repair fluency (dysfluencies %)				
Picture-story	24.30	5.95	23.10	5.89
Map	18.79	4.62	17.46	4.42
	$F = 25.936$, $p < .001$*			

Task-type repetition

Repeating the same type of task appeared to affect both accuracy and fluency. From Table 6 we can see the speakers generated more accurate language by making fewer errors on the second encounter with a similar task ($F = 10.478$, $p = .002$).

Results on fluency were more complicated, as task-type repetition had a mixed effect on the four fluency measures, with improved breakdown and repair fluency but less speed fluency (see Table 7). That is, when experiencing a similar task with different content, speakers tended to pause less frequently in the middle of clauses ($F = 9.141$, $p = .004$) and make fewer repetitions and self-corrections ($F = 4.708$, $p = .034$). However, in the meantime they seemed to speak more slowly ($F = 4.326$, $p = .041$) and produce shorter uninterrupted utterances ($F = 6.492$, $p = .013$).

Table 6. Effects of task-type repetition on accuracy

	Time 1		Time 2	
	Mean	**SD**	**Mean**	**SD**
Error-free clauses (%)	58.15	10.72	61.88	10.14
	$F = 10.478$, $p = .002$*			

Table 7. Effects of task-type repetition on fluency

	Time 1		Time 2	
	Mean	**SD**	**Mean**	**SD**
Breakdown fluency (mid-clause pauses %)	45.66	12.77	42.34	12.63
	$F = 9.141$, $p = .004$*			
Speed fluency (speech rate)	142.41	28.02	137.00	25.63
	$F = 4.326$, $p = .041$*			
Speed fluency (mean length of run)	4.99	1.25	4.69	1.03
	$F = 6.492$, $p = .013$*			
Repair fluency (dysfluencies %)	21.55	5.97	20.28	5.90
	$F = 4.708$, $p = .034$*			

Performance criteria

Table 8 shows overall that performance criteria exerted a significant impact on phrasal complexity ($F = 4.333$, $p = .018$) with the HC speakers producing the longest clauses, followed by the LC group, with the shortest clauses occurring in the CG. A further analysis revealed that only the difference between the HC and the CG reached significance ($p = .015$). Regarding subordination complexity measured by clause/AS-unit ratio, the result was less straightforward as there was an interaction between task and condition. Table 9 reveals that performance criteria were negatively associated with clause/AS-unit ratio in the picture-story task only ($F = 6.097$, $p = .006$), with the HC pairs producing significantly fewer clauses per AS-unit than their LC and CG peers ($p = .009$ and $p = .027$ respectively).

Table 8. Effects of performance criteria on phrasal complexity

	Time 1		Time 2	
	Mean	**SD**	**Mean**	**SD**
Phrasal complexity (mean length of clause)				
Control Group	5.13	0.71	5.12	0.67
Low Criterion	5.41	1.02	5.16	0.71
High Criterion	5.47	0.76	5.62	0.88
$F = 4.258$, $p = .018$* (HC>CG, $p = .015$*)				

Table 9. Effects of performance criteria on subordination complexity in picture-story task

	Time 1		Time 2	
	Mean	**SD**	**Mean**	**SD**
Subordination complexity (clause/AS-unit ratio)				
Control Group	1.26	0.19	1.21	0.13
Low Criterion	1.21	0.17	1.23	0.14
High Criterion	1.07	0.14	1.09	0.14
$F = 6.097$, $p = .006$* (HC<CG, $p = .009$*; HC<LC, $p = .027$*)				

Finally, performance criteria were also related to speed fluency measured by MLR scores ($F = 3.492$, $p = .036$), which were highest under the HC condition, followed by the CG. The scores were the lowest under the LC condition (Table 10). That is, the HC dyads produced the longest uninterrupted utterances compared with the other two groups. Differences between the highest and the lowest scores reached significance ($p = .031$).

Table 10. Effects of performance criteria on speed fluency

	Time 1		Time 2	
	Mean	**SD**	**Mean**	**SD**
Speed fluency (mean length of run)				
Control Group	5.02	1.17	4.62	0.86
Low Criterion	4.55	0.99	4.40	0.64
High Criterion	5.41	1.46	5.05	1.36
$F = 3.492$, $p = .036$* (HC>LC, $p = .031$*)				

To summarize the findings, task type, task-type repetition, and performance criteria were all associated with aspects of language output. The picture-story task gave rise

to greater syntactic complexity and lexical diversity, whereas the map task resulted in greater accuracy and fluency (breakdown and repair). A second encounter with the same type of task clearly facilitated accuracy, but led to mixed results for fluency. More specifically, repeating a similar task with different content was beneficial to breakdown fluency and repair fluency, but detrimental to speed fluency. In addition, performance criteria were overall positively related to phrasal complexity and speed fluency, but negatively associated with subordination complexity in the picture-story task.

Discussion

Task type and language production

Task type was found to be significantly related to all complexity variables with the picture-story task eliciting greater syntactic complexity and lexical diversity. This might be attributable to the differentiated discourse outcomes to which the two types of tasks gave rise. As discussed earlier, the map task, with limited communicational goals, created a context where speakers were more likely to use shorter, simpler, and less varied language, such as "turn left", "walk along the street". In contrast, participants in the picture-story task had to resort to a wider language repertoire in order to meet a wider range of functional requirements, including referring to and distinguishing between referents, following event sequences, giving and justifying personal opinions, hence resulting in more complex and diversified language use. The following are some examples:

```
Extract 1.
|um and {the, the man's} the man with a pair of glasses was {just put}
just put his right hand {on} on the {wo-} white woman's shoulder|

Extract 2.
|the boy who kick the basketball into the hole :: is try to get the
basketball out|

Extract 3.
|my picture E must be the last one :: because the boy who uh take the
basin :: {put, p-} put the water into the hole|

Extract 4.
|{uh when} uh when the long hair woman and short hair woman are talking
:: the child {make a jokes} make a joke at the short hair woman|
```

Clearly, the use of attributive adjectives, prepositional phrases and noun phrases as modifiers for referent identification (see Extracts 1) made the utterances more elaborate at the phrasal level. Extract 2 shows that relative clauses were employed to distinguish between referents. In Extract 3 *because* was used to indicate the logical relationship between clauses. The adverbial clause introduced by *when* occurring in Extract 4 suggested the sequence of events. All these give rise to subordinate clauses, resulting in more complex syntax at subordination levels.

On the other hand, the map task performers did significantly better than their picture-story task counterparts in terms of breakdown fluency (mid-clause pauses), repair fluency (dysfluencies) and accuracy (error-free clauses). Again, this is probably because the limited conceptual/communicational demands required by the map task would provide speakers with ample opportunities to use repeatedly certain syntactic structures and lexical phrases related to direction-giving. These expressions therefore are likely to be chunked in the form of multiword phrases which are stored and retrieved as a whole for later use (Kormos, 2006). The use of ready-made lexicalized chunks of this sort (also called formulaic sequences) consumes less processing capacity and functions more quickly and effectively in the ongoing communication (Skehan, 1998). As a result, learners would pause less likely in the middle of clauses to search for words or think how to verbalise what was in their mind. In the meantime, lexicalized items were rigid and memorized as a whole. Once retrieved, they were less likely to be corrected by the speakers themselves. In addition, the constant use of formulaic language might lead to fewer erroneous utterances. Boers et al. (2006) explained that formulaic sequences "constitute 'zones of safety' and appropriate use of them may thus confine the risk of 'erring' to the spaces in between the formulaic sequences in one's discourse" (p. 247).

Similar patterns with the effects of task-type difference on language performance were found in Foster and Skehan's (1996) study where they examined three types of tasks: personal, narrative, and decision-making. Their personal and narrative tasks were comparable with the map and picture-story tasks used in the present study. Foster and Skehan also reported greater accuracy and fluency in their personal task, which they ascribed to its "personal" – familiar nature. However, the present study provided an alternative account, that is, that a more limited functional requirement is what was crucial, rather than familiarity. Their narrative task (cognitively more demanding), by contrast, achieved a higher degree of complexity, seemingly at the cost of accuracy. Results of the present study also demonstrated a clear trade off between complexity on the one hand (on the picture-story task) and accuracy and fluency on the other (the map task).

Task-type repetition and language production

Generally speaking, task-type repetition was primarily related to accuracy and fluency, while syntactic complexity and lexical diversity remained unaffected. Greater accuracy was achieved when speakers were performing a new version of the same task type. This is probably because familiarity with a specific type of task may redirect part of the speakers' attention from "learning the task procedure or expressing new meaning" to form, possibly leading to improved accuracy (Mackey et al., 2007, p. 209).

However, the results were mixed with regard to fluency, where breakdown and repair fluency improved during the second encounter with a similar task, but speed fluency deteriorated. Let us first examine the improved performance in breakdown and repair fluency. As mentioned earlier, repeating a new task of the same type leads to familiarity with task procedure and helps create a context in which certain formulaic language was more likely to be activated and retrieved (Pawley & Syder, 1983, 2000). Frequent employment of formulaic language promotes effortless and smooth language production in the sense that speakers may experience fewer syntactic breakdown, and pause less frequently in the middle of clauses in an attempt to plan for the upcoming utterance. Rapid access to fixed phrases and expressions also reduces the possibility for speakers to repeat or correct themselves. In this study, a decrease in self-corrections was also possibly related to the greater use of accurate language achieved during the second task performance. Below is an extract where speakers used a number of fixed expressions when performing the map task (phrases in the identified as potential fixed expressions are underlined).

Extract 5.
B: turn right, and go along the street, on the right is the church and opposite is the office
A: yes, and go along the street, we will visit museum, museum
B: ah museum, yeah
A: yes, and turn left, turn left, we will visit mm pets
B: uh yeah, pets
A: yes, and at the end of the road we will turn right
B: yeah, [turn right

We now turn to the two speed-related measures (SR and MLR) which deteriorated during the second task performance. As speakers' language use improved on accuracy as well as on breakdown and repair fluency, due to limited processing resources, it is likely that greater attention was directed to monitoring and scarce processing capacity is left for conceptualization and formulation. At the same time, focusing on accuracy may cause difficulty in processing more language in a single encoding operation. These factors help to explain the lower SR and MLR scores observed during

the subsequent encounter with a similar task. The results concerning greater accuracy but less speed-related fluency seem to lend support to the trade-offs between performance accuracy and purely temporal fluency suggested by Brumfit (2000).

In general, these findings are consistent with the hypotheses and those reported in previous studies on the effect of task-type repetition (e.g., Bygate, 2001; Gass et al., 1999). Unlike Bygate (2001) who observed a weak effect for task-type practice on fluency, the present study, however, indicates a strong effect for task-type repetition. This is possibly because the two types of tasks employed in this study were both cognitively less demanding than those used in Bygate (2001), which may greatly reduce the processing load during the second encounter, releasing attentional resources for the improvement of language quality.

Performance criteria and language production

Statistical analyses indicate that overall performance criteria were positively related to phrasal complexity and fluency (length of run) favouring the HC group, whereas accuracy was not affected. In addition, there was an interaction between task type and performance criteria on subordination complexity. That is, task criteria were negatively associated with subordination complexity in the picture-story task, but not in the map task.

In terms of the relationship between performance criteria and phrasal complexity, the results indicated that the more demanding task criterion seemed to produce longer clauses though statistical significance was only reached for the comparison between the HC speakers and the CG. This is possibly because the need to express messages more precisely and in greater detail would push speakers to use more adjectives, adverbials, noun phrases and prepositional phrases to modify various syntactic elements. Consequently, as Norris and Ortega (2009) have pointed out, "the addition of pre- or postmodification within a phrase (via adjectives, adverbs, prepositional phrases, or nonfinite clauses)" helps to condense information, resulting in the lengthening or elaboration of finite clauses (p. 561).

The interaction between task and condition

In addition to an overall impact of performance criteria on phrasal complexity, there was an interaction between task type and performance criteria on subordination complexity. A closer look at the relationship between task criteria and syntactic complexity in the picture-story task revealed an intriguing pattern. The HC speakers produced longer clauses but fewer clauses per AS-unit, while the LC and CG generated shorter clauses but more clauses per AS-unit. This, however, was not the case with the map task. Given the three measures for syntactic complexity employed, it might first

appear that the longer the clauses or the more embedded clauses there are, the longer the AS-unit is likely to be. However, length of clauses and number of clauses per AS-unit are not necessarily linked, since a long AS-unit may consist of just one clause with a number of phrases as modifiers (Allwright et al, 1988; Ortega, 2003; Norris & Ortega, 2009). The HC speakers' preference for lengthening clauses, rather than increasing the number of embeddings to pack in more information can be explained by the one-clause-at-a-time constraint proposed by Pawley and Syder (1983, 2000), who suggested that when engaging in the transmission of a novel message, speakers can plan its syntactic and lexical content at most one single clause ahead. They also identified two main strategies adopted to create a sustained piece of discourse: the *clause-chaining* style and the *integrating* style. Native speakers frequently produce fluent units containing multiclauses, which Pawley and Syder (1983, 2000) attributed to the use of a large stock of *memorized sequences* and *lexicalized sentence stems* stored in their long-term memory. Second language speakers may be more reliant on the clause-chaining style because of limited linguistic capacity and lack of automatization in processing. This might help explain why higher phrasal complexity, but lower subordination complexity, was observed with the HC speakers in this study.

Additionally, it is plausible that much of the HC speakers' linguistic output for extensive information exchange was in the form of syntactic non-clausal units. This could happen if, due to insufficient language competence, speakers had to satisfy, under time constraints, conceptual/functional requirements that were too demanding. Resorting to phrases or fragmented language and bypassing syntax could reduce heavy processing load and ease communication pressure. It should be noted that fragmental non-clausal units were also counted as AS-units. In other words, it is very likely that a great deal of fragmented language constituted AS-units that did not contain even an independent clause. This would inevitably affect clause/AS-unit ratio in a negative way.

```
Extract 6.
A:  |{the tree, the, the ano-} the other tree {uh} uh near the hole?|
B:  |mm what?|
A:  |uh the tree in the picture F {like} like the tree in your other
picture|
```

The above extract consists of AS-units containing only a fragmental non-clausal unit, where the two speakers were negotiating the location of a tree in their picture. It is worth noting here that while such detailed information was not essential either for the primary picture-sequencing task or for the LC post-task picture retelling activity, it was of importance for the HC speakers who needed to draw the pictures.

Performance criteria were also closely associated with speed-related fluency, with the HC speakers generating a higher MLR than the other two groups, although only the difference between the HC and LC dyads reached significance. As we know, HC performers tended to pack in more information at the phrasal level by using various forms of modifiers. Conceivably, the conceptualizing and encoding of more complex semantic content would consume a greater deal of processing capacity, leaving scarce attentional resources available to attend to other aspects of language use. Hence, it would be unusual if speakers could improve fluency (MLR) simultaneously. Towell et al. (1996) proposed that changes of MLR should be interpreted by taking into account other measures of fluency. They argued that speech rate is concerned with the conceptualisation process and that MLR is concerned with formulation. Under the conditions that the proportion of speech time remains unchanged or increases, and that the length of pause does not change or decreases, runs with increased length indicate that speakers can process more language or encode more information at a single time span, hence indicating greater proceduralization of knowledge. Longer MLR may also indicate that speakers have stored prefabricated language and can have quick access to these ready-made lexical chunks. Apparently, these two accounts for increased MLR are more closely related to the development of proficiency over a longer period of time. Pawley and Syder's (1983, 2000) one-clause-at-a-time hypothesis might be more relevant here. That is, the HC speakers, in response to the task requirement, were somehow led to process more language and encode more information in a single clause, which might actually facilitate the observed higher MLR, rather hinder it. This implies then that Towell may be wrong – that differences in MLR do not reflect differences in formulation as opposed to conceptualization. In other words, differentiated conceptualization demands may also lead to differences in MLR.

The added value of the HC criterion

Overall, performance criteria had a positive impact on language complexity and fluency, but a negative influence on subordination complexity in the picture-story task. However, further examination of the source of significance revealed a clear pattern in that only differences between the HC group and either or both the other two groups reached significance. This is probably the result of a lack of discrimination concerning the perception by participants of the task requirement between the LC and the control condition. More specifically, that the additional post-task requirement stipulated by the LC condition failed to elicit different language output from that produced under the control condition is most likely either because the baseline task input was relatively simple or because the specific operationalisation of the low task criterion was not demanding enough. Hence, by means of using unfamiliar vocabulary items in the map task and requiring a minimum amount of language for each picture in the

picture-story task, speakers might have had a clearer idea of the distinction in task requirements between the LC and the control conditions, leading to differentiated language performance.

Conclusion

This study investigated the potential impact of task type, task-type repetition, and performance criteria on L2 learners' language output. It was found that tasks with limited communicational goals tended to give rise to greater accuracy and fluency, whereas tasks with more demanding functional requirements led to increased complexity. In addition, repeating a similar task with different content was likely to promote accuracy and fluency, but not complexity. These findings are consistent with results from previous studies. Further, overall tasks implemented with a high performance criterion seemed to yield greater complexity and fluency, but not accuracy. The results lend support to the limited attentional model which indicates that different dimensions of language production are unlikely to improve simultaneously.

Unavoidably there are limitations with this study. First, only two types of tasks were used in the investigation. Future studies could consider comparing language performance on more task types reflecting other discoursal features such as interviews, general discussions, and argumentations (debates), so that a fuller picture of the influence of task-type difference can be achieved. In addition, the present study reported an interaction between task type and performance criteria on subordination complexity. Using more task types would help to show whether such an interaction occurs with other task types. Second, when a similar task was repeated, increases in accuracy and some fluency measures were observed. However, based on two encounters with one task type, the present study is unable to tell us whether this pattern would carry on over subsequent performances if more exemplars of the same type were administered. Third, the study has revealed an overall positive effect of performance criterion on complexity and fluency. Unfortunately, there were no significant differences between the LC and the CG, which might be because the additional post-task requirements assigned to the LC group were not sufficiently demanding to make a difference. Future studies could consider alternative ways of operationalising task performance criteria, so as to differentiate the effects between the presence and absence of a performance criterion, as well as between the high and low performance criterion.

These findings shed some new light on the impact of task features and task implementation conditions upon second language production and provide useful implications for language teaching. A wide range of task types should be employed in the language classroom, so that learners will be able to have balanced language practice. In

addition, learners can benefit both from experiencing a similar task of the same type repeatedly and from implementing tasks with more demanding performance criteria. The latter (i.e., giving demanding performance criteria) is particularly relevant when task repetition is used. Affectively, conducting an identical task again is somewhat associated with boredom or lack of interest (Lynch & Maclean, 2001; Plough & Gass, 1993). However, with certain performance criteria pre-specified for them, learners are likely to be motivated even if they are repeating an exact same task, since they may need to re-interpret or re-define the task in some way, leading to more ambitious use of language and greater involvement in the task.

Acknowledgement

This chapter is based on the author's PhD thesis completed in August 2011 at Lancaster University, UK.

References

Allwright, R. L., Woodley, M. P., & Allwright, J. M. (1988). Investigating reformulation as a practical strategy for the teaching of academic writing. *Applied Linguistics*, 9(3), 236–256. https://doi.org/10.1093/applin/9.3.236

Appel, G., & Lantolf, J. P. (1994) Speaking as mediation: A study of L1 and L2 text recall tasks. *The Modern Language Journal*, 78, 437–452. https://doi.org/10.1111/j.1540-4781.1994.tb02062.x

Boers, F., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a Lexical Approach to the test. *Language Teaching Research*, 10(3), 245–261. https://doi.org/10.1191/1362168806lr1950a

Brown, G., & Yule, G. (1983). *Teaching the spoken language: An approach based on the analysis of conversational English*. Cambridge: Cambridge University Press.

Brumfit, C. (2000). Accuracy and fluency: The basic polarity. Reprinted in Riggenbach, H. (Ed.). (2000). *Perspectives on Fluency* (pp. 61–73). Ann Arbor, MI: The University of Michigan Press.

Bygate, M. (1996). Effects of task repetition: Appraising the developing language of learners. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching*. Oxford: Macmillan Education.

Bygate, M. (1999). Quality of language and purpose of task: Patterns of learners' language on two oral communication tasks. *Language Teaching Research*, 3(3), 185–214.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 23–48). Harlow: Pearson Education.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In J. M. L. L. B. Resnick & S. D. Teasley (Ed.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association. https://doi.org/10.1037/10096-006

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, 13, 259–294. https://doi.org/10.1207/s15516709cog1302_7

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1–39. https://doi.org/10.1016/0010-0277(86)90010-7

Coughlan, P., & Duff, P. A. (1994). Same task, different activities: Analysis of a SLA task from an activity theory perspective. In J. P. Lantolf & G. Appel (Eds.), *Vygotskian approaches to second language research* (pp. 173–193). Norwood, NJ: Ablex.

Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition*, 11, 367–383. https://doi.org/10.1017/S0272263100008391

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning*, 54(4), 655–679. https://doi.org/10.1111/j.1467-9922.2004.00282.x

Ehrlich, S., Avery, P., & Yorio, C. (1989). Discourse structure and the negotiation of comprehensible input. *Studies in Second Language Acquisition*, 11, 397–414. https://doi.org/10.1017/S027226310000841X

Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics*, 30(4), 474–509. https://doi.org/10.1093/applin/amp042

Ellis, R., & Yuan, F. (2005). The effects of careful within-task planning on oral and written task performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 167–192). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.11.11ell

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18, 299–323. https://doi.org/10.1017/S0272263100015047

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21(3), 354–375. https://doi.org/10.1093/applin/21.3.354

Gass, S. M., Mackey, A., Alvarez-Torres, M. J., & Fernandez-Garcia, M. (1999). The effects of task repetition on linguistic output. *Language Learning*, 49(4), 549–581. https://doi.org/10.1111/0023-8333.00102

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Longan, G. D. (1988). Towards an instance theory of automatization. *Psychological Review*, 95(4), 492–527. https://doi.org/10.1037/0033-295X.95.4.492

Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language teaching. *Language Teaching Research*, 4(3), 221–250.

Lynch, T., & Maclean, J. (2001). "A case of exercising": Effects of immediate task repetition on learners' performance. In M. Bygate, P. Skehan & M. Swain (Eds.), *Researching pedagogical tasks: Second language learning, teaching and testing* (pp. 141–162). Harlow: Longman.

Mackey, A., Kanganas, A. P., & Oliver, R. (2007). Task familiarity and interactional feedback in child ESL classrooms. *TESOL Quarterly*, 41(2), 285–312. https://doi.org/10.1002/j.1545-7249.2007.tb00060.x

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, 20, 83–108. https://doi.org/10.1017/S0272263198001041

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30(4), 555–578. https://doi.org/10.1093/applin/amp044

Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition*, 21, 109–148. https://doi.org/10.1017/S0272263199001047

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492–518.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–226). London: Longman.

Pawley, A., & Syder, F. H. (2000). The one-clause-at-a-time hypothesis. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 163–199). Ann Arbor, MI: The University of Michigan Press.

Plough, I., & Gass, S. M. (1993). Interlocutor and task familiarity: Effects on interactional structure. In G. Crookes & S. M. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 35–56). Clevedon: Multilingual Matters.

Robinson, P. (1995). Task complexity and second language narrative discourse. *Language Learning*, 45(1), 99–140. https://doi.org/10.1111/j.1467-1770.1995.tb00964.x

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22(1), 27–57. https://doi.org/10.1093/applin/22.1.27

Robinson, P. (2003). Attention and memory during SLA. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 631–678). Oxford: Blackwell. https://doi.org/10.1002/9780470756492.ch19

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *IRAL*, 43, 1–32. https://doi.org/10.1515/iral.2005.43.1.1

Sangarun, J. (2005). The effects of focusing on meaning and form in strategic planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 11–141). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.11.08san

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49(1), 93–120. https://doi.org/10.1111/1467-9922.00071

Skehan, P., & Foster, P. (2005). Strategic and on-line planning: The influence of surprise information and task time on second language performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 193–216). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.11.12ske

Tannen, D. (1980). A comparative analysis of oral narrative strategies: Athenian Greek and American English. In W. L. Chafe (Ed.), *The pear stories* (pp. 51–87). Norwood, NJ: Ablex.

Tarone, E. (1985). Variability in interlanguage use: A study of style-shifting in morphology and syntax. *Language Learning*, 35(3), 373–403.

Tarone, E., & Parrish, B. (1988). Task-related variation in interlanguage: The case of articles. *Language Learning*, 38(1), 21–44. https://doi.org/10.1111/j.1467-1770.1988.tb00400.x

Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, 58(2), 439–473. https://doi.org/10.1111/j.1467-9922.2008.00446.x

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–273). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.11.15tav

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, 17(1), 84–119. https://doi.org/10.1093/applin/17.1.84

Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing*, 14, 85–106. https://doi.org/10.1177/026553229701400105

Wilkes-Gibbs, D. (1997). Studying language use as collaboration. In G. K. E. Kellerman (Ed.), *Communication strategies: Psycholinguistic and sociolinguistic perspectives* (pp. 238–274). London: Longman.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, 24(1), 1–27. https://doi.org/10.1093/applin/24.1.1

CHAPTER 7

The effects of awareness-raising through stimulated recall on the repeated performance of the same task and on a new task of the same type

Chris Sheppard[1] and Rod Ellis[2,3]
[1]Waseda University, Tokyo / [2]Curtin University, Perth, Australia / [3]Shanghai International Studies University

The study explores the effects of intervening between the performance of the same task and whether these effects transfer to a new task. 40 Japanese university students completed the same monologic narrative task three times and then performed a new task of the same type (Time 4). 20 students just performed the tasks (the task-repetition group). The other 20 engaged in a stimulated recall procedure after the initial performance of the first task (the stimulated recall group). In both groups, structural complexity increased from Time 1 to Time 2 and was maintained, including in the new task. In contrast, accuracy remained the same for both groups when the same task was repeated and decreased notably in the new task. Fluency improved in both groups when the same task was repeated but declined when the new task was performed. However, performance of the new task was still significantly better than at Time 1. No differences were evident in complexity and accuracy between the task-repetition and the stimulated recall groups over time. The stimulated recall group demonstrated greater fluency than the task-repetition group when repeating the same task but not the new task. The chapter concludes with comments on the potential of stimulated recall as a means for enhancing the effects of task repetition.

Theoretical issues

Task repetition serves a similar function to pre-task planning (Ellis, 2005). That is, the first performance of a task constitutes a preparation for the subsequent performance of the same task with potential benefits on the language deployed in the second performance. Thus, research on spoken repetition has drawn on the same theoretical foundation as pre-task planning, namely Levelt's model of speech production (Levelt, 1989; Levelt, Roelofs, & Meyer, 1999).

DOI 10.1075/tblt.11.07she

When second language learners perform a task for the first time, they will have engaged in conceptualization (i.e., generated the messages they wish to communicate), formulation (i.e., accessed the lemmas and their morphological and phonological encodings from their mental lexicon), and articulation (i.e., converted the formulated message into an auditory representation). In addition, they may have engaged in self-monitoring before the point of articulation and/or after it. When they repeat the task they may benefit from the previous performance, as task repetition "can help to give learners space to work on matching meanings to language, and to integrate attention to the more redundant features (especially grammatical and phonological) into their speech" (Bygate, 2006, p. 172). More specifically, task repetition potentially aids both performance and learning in these ways:

1. Learners can reactivate the message content they have already conceptualized. Also, further online conceptualization may take place as the task is repeated resulting in changes to the original message content – for example, by simplifying, elaborating, reordering, and changing focus.
2. Learners will already have had a stab at assembling the L2 forms needed to encode the message content. Even if the task is repeated after a substantial time gap there are likely to be traces of the initial form-meaning mappings in their long-term memory which can therefore be accessed more readily than on the first occasion.
3. Similarly, it is reasonable to suppose that because learners have already articulated the formulated message once, they will more easily engage with the phonological and phonetic encoding that this entails and the relevant motor activity.
4. Finally, repeating the task may also have an effect on the extent and nature of the monitoring that occurs. Familiarity with the task may lead to less extensive monitoring or it might lead to increased monitoring if learners focus more on grammatical accuracy and draw on their declarative knowledge of target rules.

These potential advantages of task repetition can be explained in terms of the nature and role of working memory in L2 production. While there are differences in models of working memory (e.g., Baddeley, 2007; Miyake & Shah, 1999), they all acknowledge that it is limited in capacity (i.e., it can only hold a certain amount of information at a time) and that therefore there are constraints on what learners can attend to. Skehan (1998) suggested that there will be competition between attending to meaning and form and, in the case of form, between complexity and accuracy. His research on the effects of pre-task planning (e.g., Skehan & Foster, 1999) led him to propose that it serves to relieve pressure on online planning when a task is performed as learners will already have conceptualized the message content. This facilitates fluency and may also allow learners to attend more fully to form. Similarly, task repetition can be expected to lead to improved fluency and may also allow for selective attention to be focussed more easily on form.

There are, however, some important questions that need to be asked. If there is competition between the two dimensions of form – complexity and accuracy – what aspect does task repetition benefit? While there are studies that show pre-task activity assists both complexity and accuracy (thus possibly contradicting Skehan's Trade-off Hypothesis) there are also several studies that show it benefits either complexity or accuracy but not both (see Ellis, 2009). Does task repetition have similar effects to pre-task planning?

There is also a more important question. Arguably, for task repetition to become accepted as a pedagogic strategy it will be necessary to show that it can lead to measurable gains in the use of the L2. To investigate this it is necessary to examine whether repeating tasks helps learners to perform new tasks more fluently and/or more accurately and with more complex language. There are theoretical grounds for believing that it might. The Transfer Appropriate Processing Hypothesis (Lightbown, 2008) predicts that there is a direct relationship between the type of practice that the learner experiences and the kind of learning that takes place. Thus we might expect that the benefits from repeating a task will transfer to the performance of a new task especially if there are points of commonality between the old and new task.

Previous studies of task repetition

The research has investigated different types of task repetition:

1. Procedural repetition – "refers to carrying out several tasks that require the same procedure to accomplish the communicative goal but require different content knowledge".
2. Content repetition – "refers to carrying out the same tasks that require different procedures to accomplish the communicative goal but require the same content knowledge".
3. Task repetition – "refers to carrying out the same task repeatedly, using the same procedure to accomplish the communicative goal and requiring the same content knowledge". (Patanasorn, 2010, p. 13)

The study we will report considered only task repetition, which is the type that has received the most attention from researchers.

Manchón (2014) distinguished five types of repetition studies, two of which are relevant to the present study:

1. Studies that have investigated the effects of one or more types of task repetition on the characteristics and features of oral production (e.g., Bygate, 2001; Gass,

Mackey, Alvarez-Torres, & Fernandez-Garcia, 1999; Kim & Tracy-Ventura, 2013; Lynch & McLean, 2000; Patanasorn, 2010).

2. Studies that have investigated the effect of introducing some kind of intervention between the first and subsequent performance(s) of a task (e.g., Baleghizadeh & Derakhshesh, 2012; Hawkes, 2011).

In task repetition studies, researchers investigate the effect of task repetition on the performance of the same task or type of task (e.g., Bygate, 1996; Birjandi & Ahangari, 2008) and sometimes on a new task (e.g., Kim & Tracy-Ventura, 2013; Patanasorn, 2010). This distinction is important as it relates to whether task repetition has an effect on *performance* or, more importantly, whether transfer to new tasks occurs, indicating an effect on L2 *development*.

The studies vary in a number of potentially significant ways, although this variation has not been systematically investigated. One source of potential variation is whether the task to be repeated involves monologic discourse (e.g., Birjandi & Ahangari, 2008; Gass et al., 1999) or dialogic discourse (e.g., Kim & Tracy-Ventura, 2013; Lynch & Mclean, 2000; Patanasorn, 2010). Tasks also vary in terms of the type of discourse they elicit – narrative in many of the studies (e.g., Bygate, 2001; Gass et al., 1999) but also descriptive (e.g., Kim, 2013), interview (e.g., Bygate, 2001), and argumentative (e.g., Birjandi & Ahangari, 2008). As Fukuta (2016) noted, other variables include the number of repetitions – for example, just once in Gass et al. (1999) and in Bygate (2001)[1] but three times in Kim (2013). The length of time between repetitions also varies – for example, in Kim (2013) it was one day, in Gass et al. (1999) 1 week, and in Bygate (2001) 10 weeks. The setting differs as well: some of the studies were carried out in real-life classrooms (e.g., Kim, 2013) while others (e.g., Birjandi & Ahangari, 2008; Gass et al., 1999) were laboratory-based.

Finally there are differences in how the effects of task repetition have been measured. In some studies global measures were used. Gass et al. (1999), for example, used holistic ratings. Other studies (including Gass et al.) focused on specific linguistic features, typically grammatical, for example the use of past tense in narratives (e.g., Kim, 2013; Patanasorn, 2010). In general, however, reflecting how task-based performance has been typically measured, researchers chose to evaluate the effects of task repetition in terms of complexity, accuracy, and fluency, although how these constructs were measured differs considerably.

Clearly the disparity in the design of the studies makes it difficult to reach any clear conclusions about the effects of task repetition on either performance or devel-

1. In Bygate (2001), the learners also repeated different tasks in the 10 weeks between the two tasks that were analysed for complexity, accuracy and fluency.

opment.[2] A crude but widely accepted generalization is that task repetition benefits fluency and sometimes complexity but is less likely to lead to gains in global accuracy, although it may result in improved accuracy of task-relevant target features (e.g., past tense in story-telling tasks) in the repeated task. By and large, task repetition appears to have negligible effects on the language used in a new task. Ellis (2009) suggested that this may be because in most of the studies there was no intervention directing the learners' attention to language.

Task repetition studies involving intervention

Lynch (this book) discusses a number of ways of conducting an intervention between performances of a task – the provision of feedback, preparing and correcting a transcript of the first performance, listening to or watching a comparator recording of a native speaker or non-native speaker performing the task. Lynch provides a thoughtful summary of his own previous classroom-based research involving various kinds of intervention, concluding that "the various forms of enhancement built into the tasks were successful in creating a context in which learners did not perceive the second round of performance as a mere 'verbatim repetition'" (p. 223).

There have, however, been very few experimental studies investigating the effects of intervention. Baleghizadeh and Derakhshesh (2012) reported a small-scale study of four intermediate Iranian adult EFL learners. The learners first read chapters in a graded reader and then gave an oral presentation to the class summarising the plot, which was audio recorded. They then transcribed the recordings, correcting any errors they noticed. A researcher made further corrections and also provided some metalinguistic explanations. The corrected transcripts were given back to the learners who then repeated their oral presentation. The intervention led to the learners fixing nearly half of their errors when they repeated the task.

In another small-scale study, Hawkes (2011) investigated the performances of the same tasks by six pairs of 2nd year junior high school students in Japan. Between performances of the tasks they received direct instruction followed by practice of predetermined structures and vocabulary relevant to each task. Results showed an increase in the number of lexical and grammatical corrections for all the tasks in the repeat performance and in the number of pronunciation corrections for one of the tasks. However, the results for individual dyads showed that the increase in corrections was quite limited.

2. Task is inevitably a cluster of features. One of the continuing problems with quantitative studies of tasks is the difficulty of isolating the effects of specific design and implementation variables. See Ellis (2015).

Three studies examined the effect of another type of intervention – manipulating the time available when learners repeated a task. Ahmadian and Tavakoli (2011) reported that the provision of online planning time improved accuracy (replicating Yuan & Ellis, 2003) but there was no additional advantage for task repetition. Also, careful online planning along with task repetition enhanced fluency and led to more complex language. Two other studies tested the claims of Nation and Newton's (2009) 4/3/2 technique. This involves gradually increasing the time pressure as learners repeat the task, thereby reducing the time available for online planning. De Jong and Perfetti (2011) compared two groups of L2 learners, one of which repeated the same task (exact repetition) three times and the other three tasks which had different topics but the same procedures (procedural repetition). They reported that fluency improved in both groups over time but was only maintained in post-tests in the exact repetition group. Thai and Boers (2016) found that increasing time pressure led to more fluent production but did not result in more complex or accurate language. In contrast, those learners who performed the task three times without any increase in time pressure manifested small gains in fluency but some gains in complexity and accuracy. Thai and Boers concluded that time-pressured task repetition is not a useful pedagogic procedure and that teachers would benefit more from providing "opportunities for language adjustment early on in the task sequence" (p. 19) (but see de Jong & Tillman, this volume, for an alternative conclusion).

These few studies suggest that when learners' attention is focused on form either by means of some intervention between performances of a task (as in Baleghizadeh & Derakhshesh's and Hawkes' studies) or by allowing time for careful planning when they repeat the task (as in Ahmadian & Tavakoli) improvements in the quality of language produced occur. The study reported in this chapter extends the research by comparing the effects of task repetition on complexity, accuracy, and fluency in both the same task and a new task in groups that (1) repeated the tasks without any time pressure or focus on form intervention and (2) repeated the tasks without time pressure but with a between-task intervention involving stimulated recall.

Stimulated recall

Stimulated recall is an introspective method that is used to prompt participants to recall thoughts that they had while performing a task (Gass & Mackey, 2000). It involves presenting learners with audio or video recordings of a task and asking them to comment retrospectively on their performance by revealing the thought processes they engaged in at the time. Stimulated recall has been widely used in SLA research, often as a means of obtaining evidence of what learners 'noticed' in the input or attended to in their own output. In our study, stimulated recall is used as a pedagogical procedure

to induce learners' awareness of how they performed a task and thus, potentially, to promote learning.

Previous research has shown that stimulated recall is characterized by reactivity. Egi (2008) pointed out that stimulated recall can be reactive in two ways. First, it serves as additional input when learners are presented with audio or video recordings of their own performance of a task. Second, the act of verbalisation may have an effect on their awareness. Thus while stimulated recall remains controversial as a research tool, it might prove valuable as a form of intervention. Such a view is supported by sociocultural theory. Swain (2006) drew on Smagorinsky (2001) to argue that "the process of rendering thinking into speech is not simply a matter of memory retrieval, but a process through which thinking reaches a new level of articulation" (171–3) and that because "thinking becomes rearticulated through the process of speech," it is "an agent in the production of meaning" (p. 240).

Some evidence for Swain's argument comes from Adams (2003). The learners in her study first wrote a text and then compared the text to a reformulated version. Some of the learners were then asked to recall their thoughts about the differences between their own and the reformulated text. The learners who took part in the stimulated recall achieved higher accuracy scores than those who did not. Egi's (2007) study also indicated that stimulated recall has an effect on learning. However, in another study Egi (2008) failed to find any evidence of reactivity. A group of learners who completed a stimulated recall activity following the performance of communicative tasks designed to elicit the use of specific Japanese grammatical features did no better on either an immediate or delayed post-test than a group that just performed the tasks. These mixed results may reflect the fact that, in Adams and Egi (2007), the stimulated recall was carried out one-on-one in a laboratory type setting, but in Egi (2008) it was a classroom setting. It is possible that the depth of reflection elicited by stimulated recall is greater when it takes place face-to-face with a researcher than with a teacher in a whole-class environment.[3]

In Lindgren and Sullivan (2003), stimulated recall was seen as a trigger for inducing noticing. In their study, two young Swedish learners of English completed two computer-mediated writing tasks, the first without stimulated recall and the second with. Lindgren and Sullivan reported that both learners added more words to their revised text and also more typographical text revisions involving spelling and grammar after stimulated recall. Similar results were obtained by Fukuta (2016) for L2 oral production. In this study, relatively high-proficiency L2 learners who took part in a stimulated recall session following the first performance of a narrative task demon-

3. The learners in Egi (2007) also received training in stimulated recall while this does not appear to have been the case in Egi (2008).

strated gains in accuracy and lexical variety when the task was repeated one week later but no gains in fluency or complexity.

These studies afford quite mixed results regarding the impact of stimulated recall on the repeated performance of a task and point to the need for further research. Potentially, stimulated recall can bring a breadth of focus to how students perceive a task and the meanings they wish or need to convey. It has the potential to attract their attention to different aspects of conceptualization, formulation, and articulation and thus impact on the complexity, accuracy, and fluency of their second performance of a task. In accordance with Skehan's limited resources model, however, it is unlikely that stimulated recall will result in enhanced repetition in all three aspects of language use. As Fukuta found, students are likely to elect to make use of it to attend to just one aspect.

In the study reported below we used stimulated recall to address the following research questions:

1. What effect does participating in a stimulated recall session, following the first performance of an oral narrative task, have on the complexity, accuracy, and fluency of L2 learners' production when they repeat the same task?
2. What effect does participating in a stimulated recall session, following the performance of an oral narrative task, have on the complexity, accuracy, and fluency of L2 learners' production when they perform a task of the same type but with different content?

These questions were informed by the research referred to above, which suggests that raising learners' awareness of how they performed a task through stimulated recall can impact on their subsequent performance of the same task and on learning.

Method

Participants

The data for this experiment were collected from 40 Japanese university students recruited from elective beginning English courses at a language institute attached to a large private university in Tokyo. They agreed to participate after receiving an explanation of the research project and were offered 2000 yen in return for their time.

The students ranged from first year undergraduates to second year master's students, with the majority enrolled in the first three years of undergraduate study, and they came from a large number of different majors. There were 17 males and 23 females. The range of their ages was 18–23 years with one participant in his fifties. The students did not report having spent any substantial time overseas for the purposes of

English study. These students can be characterized as false-beginners in the Japanese context, having studied English as a foreign language for at least 6 years at high-school. Most of them had passed a rigorous English examination to be admitted into the high-prestige university. However, speaking, which is the focus of this study, was not tested in this exam, and, accordingly, was not the focus of their six years of education. The participants were beginning-level speakers.

Design

All the participants completed the same one-way narrative task three times (Times 1, 2, and 3) and then performed a new task of the same type (Time 4) to investigate whether there was any transfer effect of the task repetition. The participants were assigned randomly to one of two groups. One group (the Task Repetition Group) just completed the tasks. The other group (the Stimulated Recall Group) took part individually in a stimulated recall (SR) session immediately after they had completed the task the first time. The SR session involved participants viewing a video recording which was paused whenever repetition or reformulation occurred and then being asked to say what they were thinking about.

Instruments

Two monologic narrative tasks in the form of picture stories were used to elicit repeated oral performances. Given the false beginner status of the participants, the material was chosen so as to try to ensure that the task conditions were not too demanding (Ellis, 2003). The first task consisted of an incorrectly-ordered, 10-frame picture story which was visible to the participants while they were speaking, making the task 'here-and-now' (see Robinson, 1995). The topic of the task was a family dinner, which was selected as it was well within the participants' experience. The task required the participants to provide links between the events shown in the pictures. The researcher acted as the listener, putting the jumbled set of pictures depicting the story into the correct order to achieve the task outcome.

The new task involved new content (a dinner at a restaurant with friends) but was performed in the same way as the first (i.e., it involved procedural repetition). In order to confirm that these two tasks were equivalent, a pilot study was conducted with 21 participants drawn from the same population but from different academic years. The pilot study confirmed that for fluency ($F(1,21) = 2.314$, $p = .143$) and accuracy ($F(1,21) = .065$, $p = .801$), the tasks were equivalent. Complexity, however, was found to be significantly greater in the first task ($F(1,21) = 4.523$, $p = .045$) with a medium

effect size (Cohen's $d = 0.657$).[4] However, this does not compromise the study since any significant increase in complexity from first to second task can be judged to be a real improvement.

Experimental procedure

Each participant came to the researcher's office and completed the tasks individually. The participants were randomly assigned to one of the two groups while ensuring that each of the groups had equal membership. The participants in both groups were given one minute preparation time and then performed Task One for the first time. This task took an average of 4.40 minutes (*SD* = 1.89 *min*) to complete. Production was recorded by a video camera and a digital audio recorder.

Following the first performance of the task, the participants in the Task Repetition Group (who would not be doing the SR procedure) took part individually in English conversation with the researcher for roughly the same length of time as the SR procedure would take – as determined in a pilot study. The conversations were mainly focused on the participants' daily lives and on university activities, and lasted on average 28.5 minutes (*SD* = 1.96 *min*).

The SR group read the procedure instructions in Japanese. The video tape of each participant's performance of the narrative was then replayed. The participants also had access to the picture story. It was hoped that the two aids would provide the stimulus necessary to assist recall of the thought processes they had had while performing the task. During the video presentation, the researcher paused the video whenever there was a silent pause, a filler, a repetition, or a reformulation in the student's production and then stated in Japanese:

> "You paused after you said _____", in the case of a silent pause, or "You said _____," in the other cases. "What were you thinking at this time?"

Care was taken not to deviate from this wording.

The participants were also informed that they could make comments at any time during the procedure. If they started to speak, or it seemed like they wanted to speak, the video was paused. The video tape was also replayed as many times as necessary to assist their recall. Although the participants were given a choice of whether to conduct the sessions in English or Japanese, all chose Japanese. Examples of the SR procedure are provided below.

---

4. The Stimulated Recall Group did gain in complexity at Time 2 but the difference between Time 1 and Time 2 did not reach significance because the variance (*SD*) was large – much larger than that of the Task Repetition Group at Time 2 (see Table 1).

Example 1.
Trigger: They are, they, ah, next picture, they look so happy
R： they are と言った後、あ、next picture と言った。なにを考えていまし たか？
M2(2)： 私の中では次の絵に移ったけれども、いきなりthey are と言ったら、聞いている人が同じ絵の中で動いていると思ってしまうかなあと 思って、一応区切った
[R: First you said 'they are', and then after that you said 'next picture'. What were you thinking?]
[M2(2): I was already thinking about the next picture, but I thought that if I suddenly said 'they are', then the person listening would think that I was talking about the same picture, so I changed what I was saying.]

Example 2.
Trigger: Ah, second picture, ah, the man, ah, asked the father, ah, to, ah, where Kate is
R： ask の後にポーズがあった。その時、なにを考えていましたか？
T1(2)： Ed が「Kate がどこにいるのか」と聞いているというのをまず日本語で考えて、それを英語に置き換えようとしてポーズができた。
[R: There was a pause after 'ask'. What were you thinking at that time?
[T1(2): I first made the sentence 'Ed asked where Kate was' in Japanese. Then during the pause, I was changing the sentence into English.]

These sessions were recorded by Digital recorder. They took 25.7 minutes to complete on average. The standard deviation of the sessions (7.11 *min*) was quite large compared to the conversations. One of the reasons for the large variation was the length of the individual participants' narratives at Time 1 which accounted for 22.3% of the variance ($r = .473$, $p < .01$).

Following the stimulated recall treatment, the participants were asked to repeat the same narrative picture story task (Time 2). The participants were again given one-minute preparation time before they spoke. The mean completion time for the task repetition was 3.89 minutes ($SD = 1.67$ *min*). The task repetition was recorded with the digital audio recorder but this time it was not video recorded.

The second task repetition (Time 3) and the performance of the new task (Time 4) took place two weeks after the first repetition (Time 2) – a time gap similar in length to previous studies (e.g., Gass et al., 1999; Bygate, 2001). Participants were asked to perform the new task (Time 4) directly after the second task repetition (Time 3) to reduce the possible impact of extraneous factors and to maximize the impact of the stimulated recall procedure on learners' awareness of how they performed the new

task. One-minute preparation time was again provided before the performance of each task. The second task repetition took an average of 3.61 minutes (*SD* = 1.66 *min*). Finally, the new task, took longer to complete (*M* = 5.56 *min*, *SD* = 2.97 *min*).

Analysis

The audio recordings were transcribed broadly by the researcher but with all fillers, repetitions, and formulations included. The transcripts were then analysed for the complexity, accuracy, and fluency of the participants' production. The measures were selected following an exploratory factor analysis of 29 measures (18 fluency measures, 7 complexity measures, 4 accuracy measures). The result of the analysis suggested a construct validity of the complexity, accuracy, and fluency measures and provided the basis for selecting the specific measures used in the study, which with one exception were those with the highest loadings. Grammatical complexity was measured by phrases per t-unit, which had a .911 loading on the complexity factor. Accuracy was operationalized as percentage of error free clauses, which had a loading of .915 on accuracy. Fluency was measured in terms of pruned words per minute (*loading on fluency* = .876). The highest loading measure for speaking rate was syllables per minute (.938) but the measure pruned words per minute was chosen because it was the only fluency measure that did not load on some other factor. For a full description, see Sheppard (2006).

The transcripts and the sound files for each task performance were analysed as follows:

1. Words produced by each participant were counted. Repetitions, reformulations and fillers were subtracted from this figure. (*Inter-rater reliability* = .995).[5]
2. T-units were counted following Hunt (1966). This measure was designed for written samples but is also valid for monologic oral samples. To deal with the disfluencies which were present, repetitions and fillers were ignored, and fragmented sentences were either attached to another unit or counted separately based on the context. (*Inter-rater reliability* = .93).
3. Clauses, including independent, sub- and embedded clauses were also counted. (*Inter-rater reliability* = .96)
4. The number of phrases was counted by tallying adverbial phrases (place, time, manner etc.), subject-, and object-noun phrases, and verb phrases (see Quirk, Greenbaum, Leech, & Svartvik, 1985, Section 2.3). (*Inter-rater reliability* = .59)
5. Error free clauses were determined. (*Inter-rater reliability* = .68)

5. Interrater reliability was confirmed by Pearson's correlations correlating the researchers scores with that of a specialist in the field of second language education on 9.75% of the data.

6. The total length of each performance of a task was determined in minutes. (*Inter-rater reliability* = .99)

The CAF scores were then calculated using these measures to create the appropriate ratios listed above. These were entered into a 2 X 4 mixed design ANOVA analysis. The between-subject variable was group – the Task Repetition Group and the Stimulated Recall Group. The within-subject measure was time: the performances of the tasks at Time 1, Time 2, Time 3, and Time 4. Greenhouse & Geisser's (1959) three step procedure was used to account for possible sphericity violations and the degrees of freedom were corrected using Greenhouse-Geisser's ε. Significance was set at $p < .05$. Pair-wise post-hoc comparisons were undertaken and *p*-values were adjusted using Tukey's HSD procedure.

Results

The descriptive statistics for the complexity in the students' production are shown in Table 1 and Figure 1. The mixed design ANOVA shows that there is a significant interaction for Group and Time (Greenhouse-Geisser's $\varepsilon = .87$, $F(2.61,99.2) = 2.80$, $p = .043$, $\eta_p^2 = 0.022$). However, there is no significant main effect for either Group ($F(1,114) = 0.064$, $p = .800$, $\eta_p^2 = 0.0012$) or Time ($F(3,114) = 2.24$, $p = .087$, $\eta_p^2 = 0.017$). The post hoc tests show that the Group-Time interaction effect is due to group differences between Time 1 (first performance of Task One) and Time 3 (the second repetition of the task) ($F(1,38) = 9.83$, $p = .003$, $\eta_p^2 = 0.042$). The number of phrases per T-unit increased for the Task Repetition Group in comparison to the Stimulated Recall Group although the mean increase was quite small (i.e., only 0.13 phrases per t-unit).

Table 1. Phrases per T-Unit over time by group

Time	**Group**			
	Task Repetition		**Stimulated Recall**	
	Mean	***SD***	**Mean**	***SD***
Time 1	1.26	0.13	1.33	0.24
Time 2	1.33	0.14	1.40	0.33
Time 3	1.39	0.14	1.32	0.20
Time 4 (New Task)	1.36	0.18	1.34	0.25

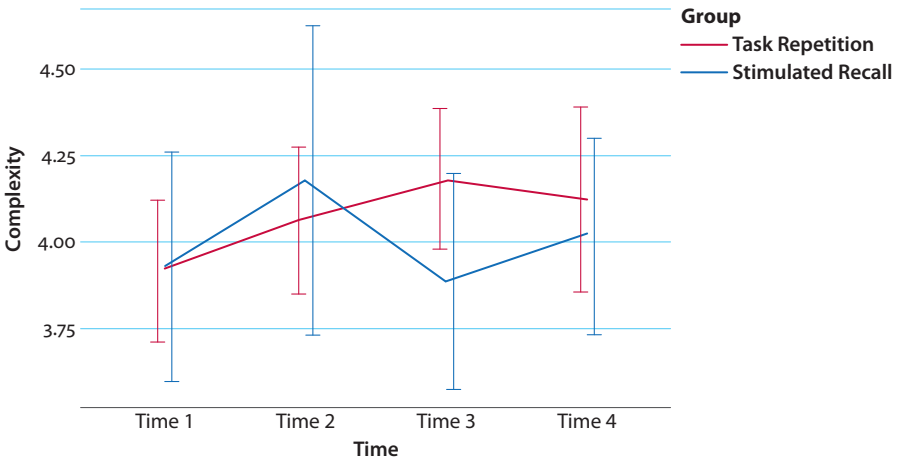

Figure 1. Phrases per T-Unit over time by group

The descriptive statistics for error free clauses are shown in Table 2 and Figure 2. The mixed design ANOVA shows that while there is no interaction effect between group and time ($F(3,114) = 0.407$, $p = .749$, $\eta_p^2 = 0.007$), or an effect for group ($F(1,114) = 0.670$, $p = .412$, $\eta_p^2 = 0.007$), there is a significant effect for time (Greenhouse-Geisser's $\varepsilon = .87$ $F(2.61,99.2) = 7.76$, $p < .001$, $\eta_p^2 = 0.113$). The post-hoc test shows that both groups had a lower ratio of error free clauses in the New Task ($F(1,38) = 8.98$, $p = .005$, $\eta_p^2 = 0.115$) compared to Time 1 (first performance of Task One).

Table 2. Error free clauses over time by group

Time	Group			
	Task Repetition		Stimulated Recall	
	Mean	*SD*	Mean	*SD*
Time 1	0.35	0.11	0.38	0.17
Time 2	0.36	0.14	0.42	0.17
Time 3	0.38	0.13	0.38	0.13
Time 4 (New Task)	0.26	0.14	0.27	0.13

The third measure included in this study was the number of pruned words per minute. The descriptive statistics are shown in Table 3 and depicted in Figure 3. Significant differences were found for time (Greenhouse-Geisser's $\varepsilon = .94$, $F(2.61,99.2) = 11.77$, $p < .001$, $\eta_p^2 = 0.086$), and for the time/group interaction (Greenhouse-Geisser's $\varepsilon = .94$, $F(2.61,99.2) = 3.97$, $p = .010$, $\eta_p^2 = 0.031$). There was no significant group difference ($F(1,114) = 1.20$, $p = .281$, $\eta_p^2 = 0.021$). The significant effect for time was due to the fact that the number of pruned words per minute increased for both groups from Time 1 to all subsequent task performances: Time 1 vs Time 2 ($F(1,38) = 31.59$,

$p < .001$, $\eta_p^2 = 0.121$), Time 1 vs Time 3 ($F(1,38) = 25.06$, $p < .001$, $\eta_p^2 = 0.120$), and Time 1 vs Time 4 (the New Task) ($F(1,38) = 5.52$, $p = .024$, $\eta_p^2 = 0.039$). The post-hoc tests for the time/group interaction revealed that the differences lay between Time 1 and Time 2 ($F(1,38) = 6.96$, $p = .012$, $\eta_p^2 = 0.029$). These pairwise comparisons indicated it was the Stimulated Recall Group that produced more pruned words per minute from Time 1 to Time 2 than the Repetition Group. There was also a significant difference between Time 3 and Time 4 (the New Task) ($F = (1,38) = 6.21$, $p = .017$ $\eta_p^2 = 0.029$), which reflects the lower order interaction between group and time between these two times ($F = (1,38) = 5.49$, $p = .024$, $\eta_p^2 = 0.026$). The drop is due to the loss of fluency between Time 3 and Time 4 in the Stimulated Recall Group. The Repetition Group demonstrated no significant change over these times.

Table 3. Pruned words per minute over time by group

Time	Group			
	Task Repetition		Stimulated Recall	
	Mean	*SD*	Mean	*SD*
Time 1	36.2	15.2	37.0	14.5
Time 2	42.2	11.4	53.7	20.3
Time 3	43.8	14.1	51.4	17.1
Time 4 (New Task)	43.5	14.2	41.4	15.4

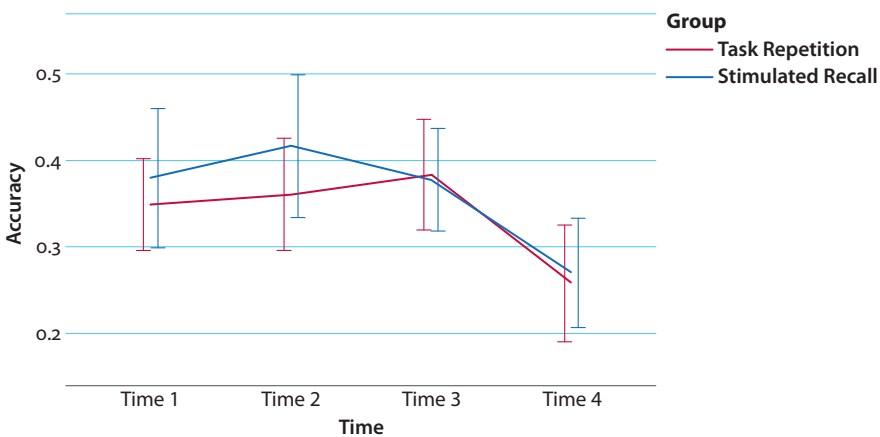

Figure 2. Proportion of error free clauses over time by group

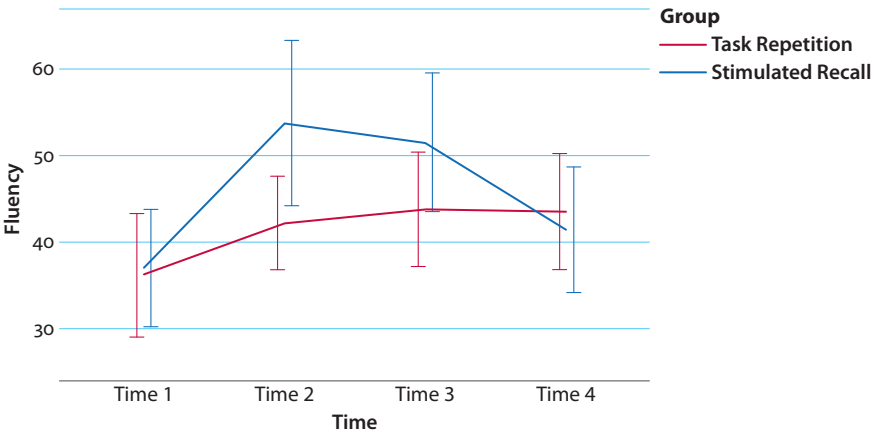

Figure 3. Pruned words per minute over time by group

Summary

The results for the effects of the stimulated recall on the performance of the same task were as follows:

1. Complexity
   There was no effect for stimulated recall on complexity; in fact, only the Task Repetition Group produced significantly more phrases per t-unit at Time 3 than at Time 1 although the effect was small, as is evident in the group X time interaction ($F(1,38) = 9.83$, $p = .003$, $\eta_p^2 = 0.042$).
2. Accuracy
   There was no effect on accuracy; there were no significant differences found in the repeated performances of the Stimulated Recall Group and Task Repetition Group and no difference between the two groups either.
3. Fluency
   There was an effect on fluency; both groups improved in fluency from Time 1 to Time 2 but the Stimulated Recall Group produced significantly more pruned words per minute at Time 2 compared to Time 1 than the Repetition Group, although the effect was small ($\eta_p^2 = .029$). However, the interaction between Group and Time (Time 1 and Time 3) was no longer significant in the second repetition, with the effect size dropping by more than half ($\eta_p^2 = .013$).

The results for the effects of the stimulated recall on the performance of the new task were as follows:

1. Complexity
   No effect was found. There was no interaction found between Group and Time for complexity in relation to the New Task (Time 4).
2. Accuracy
   Both groups demonstrated lower accuracy (i.e., fewer error free clauses) in the New Task than in the first performance of Task 1 with a medium effect size for both groups (d = .715 and d = .767 for the Task Repetition and Stimulated Recall Groups respectively).
3. Fluency
   There was a significant increase in speech rate for both groups from T1 to T4 (the performance of the new task), with a medium effect size for the Task Repetition Group (d = 0.50) and a small effect size for the Stimulated Recall Group (d = 0.29). However, the Task Repetition group maintained its level of fluency from Time 3 to Time 4 (the New Task) whereas the Stimulated Recall Group lost previous gains in fluency. Falling to the same level as that of the Task Repetition Group.

Discussion

Research question 1 asked whether the stimulated recall added anything to the effect of task repetition. The answer is that it had little additional effect.[6] Both task repetition and stimulated recall + task repetition led to greater fluency. In the short term, the stimulated recall had a greater effect but this was not maintained over time. Otherwise the effect of the stimulated recall appears to have been limited, as, unlike plain task repetition, it did not enhance complexity. Also, whereas the Task Repetition Group manifested small (but non-significant) gains between Time 2 and Time 3, the Stimulated Recall Group manifested small (but non-significant) losses in all three aspects of language use.

Research question 2 asked whether there was any difference in the complexity, accuracy, and fluency of the two groups' performance of the new task. As seen in Figures 1 and 2 the two groups' scores for all three dimensions were very similar. In other words, the initial advantage for fluency reported for the Stimulated Recall Group at Time 2 was not only not maintained in the second repetition of the original task (i.e., at Time 3) but also did not transfer to the performance of the new task at Time 4. There was a steady decline in fluency in this group from Time 2 onwards. In

6. It is also entirely possible that the measures employed in this study were too coarse to uncover small improvements in production due to awareness-raising. It would be prohibitively difficult to uncover such improvements statistically as each participant is likely to have focused on a different aspect of his performance.

contrast, the task Repetition Group maintained its fluency from Time 3 to Time 4. Previous research (Bygate, 2001; Gass et al., 1999; Patanasorn, 2010) has shown that the effects of repeating the same task may not transfer to a new task. In this study, the Task Repetition Group did achieve a higher level of complexity in the new task than in the first performance of the first task but the difference was not statistically significant.[7] The Stimulated Recall Group's new narrative manifested almost the same level of complexity as the first performance of the original task. Accuracy declined significantly for both groups in the new task. The clearest evidence of transfer of effect to a new task can be seen in fluency. Both groups were more fluent in the new task (i.e., at Time 4) than at Time 1.

These results largely support the findings of a number of previous studies of exact task repetition (e.g., Ahmadian & Tavakoli, 2011; Gass et al., 1999; Patanasorn, 2010), namely, repeating the same narrative task results in greater complexity and fluency but has little effect on global accuracy. It would seem then that performing a task aids both conceptualization and formulation so that when the task is repeated, low-proficiency learners – such as those in this study – can more readily access the language they need to encode the content, resulting in increased fluency, and also use the savings in working memory to expand their utterances to include more phrases per T-unit. The lack of any effect for task repetition on global accuracy can be explained by the nature of the task, which prioritized meaning over form, and by Skehan's (1998) Trade-off Hypothesis, which posits competition between fluency, complexity, and accuracy. There is, however, the question of why trade-off effects continue to be evident when a task is repeated. Perhaps, limited-proficiency learners such as those in this study require more than two or three repetitions to achieve sufficient control of the language for performing the task to enable them to switch attention from fluency and complexity to accuracy.

Unlike in Fukuta (2016), the stimulated recall did not lead to greater accuracy. Like the Task Repetition Group, the Stimulated Recall Group gained in fluency but no statistically significant gains were evident in either complexity or accuracy.[8] Furthermore, as Figures 1, 2, and 3 show, any beneficial effect for the stimulated recall intervention tailed off when the task was repeated again at Time 3. We speculate, therefore, that the overall effect of the stimulated recall was to focus learners' attention

7. During a pilot study, the first task was found to elicit greater complexity than the new task. It is important to note here that there was no difference between the groups in complexity. However, we cannot entirely discount the possibility that entirely equivalent tasks may have produced a significant transfer effect for complexity.

8. There are other possible explanations for why this study failed to find any effect for the stimulated recall intervention on accuracy (e.g., differences in the learners' proficiency and the length of time between task repetitions)

even more fully on the content of the story, in line with Swain's (2006) claim that rearticulated thinking serves as an agent of meaning and enables speakers to reach a new level of articulation. As Egi (2008) noted, stimulated recall provides learners with input and, as carried out in this study, it drew attention to the learners' dysfluencies. Enhanced familiarity with the story's content together with a heightened awareness of their dysfluencies may have encouraged the learners to aim for a smooth rendition of the story. Disappointingly, however, this effect did not carry over to the final performance of the same task at Time 3 (two weeks later).

Thus, in comparison to previous studies – Adams (2003), Egi (2007), Lindgren and Sullivan (2003) and Fukuta (2016) – the stimulated recall had no clear effect on linguistic selection. The explanation probably lies in differences in how the stimulated recall was carried out in this and the other studies. In this study it focused learners' attention on their filled and unfilled pauses and elicited comments related to conceptualization. Several learners commented that they initially formulated what they wanted to say in Japanese. Perhaps in our study the stimulated recall led to learners engaging in an internal repetition, as suggested by Bygate (1996), rather than formulation, whereas in the other studies it directed their attention to the choice of language. Intervention that focuses explicitly on learners' linguistic errors (e.g., Baleghizadeh & Derakhshesh, 2012; Fukuta, 2016) has been found to have a notable effect on accuracy when the task is repeated.

As with all research, the measures used limit the conclusions which can be made from this study. The accuracy measure (error-free clauses per clauses) used in this study, for example, is global. It cannot reveal small improvements in the accuracy of specific linguistic features. Finer-grained measures would be needed to examine the efficacy of stimulated recall on particular linguistic features.

Overall, then, the results of this study support those of previous studies that have reported an effect for repeating a task on fluency and complexity but no effect for accuracy. Even though the stimulated recall had little effect, we believe that it constitutes a promising type of intervention between repetitions of a task. We suggest that for stimulated recall to impact on complexity and accuracy, it would be necessary to direct learners' attention quite explicitly to the choice of linguistic forms as in Fukuta's (2016) study.

Conclusion

Stimulated recall should be seen as a variable procedure. In our study, it focused on learners' pauses and was found to have a short-term effect in enhancing fluency but no effect on complexity or accuracy. In other studies (e.g., Fukuta, 2016) it has focused on learners' choice of language and has been found to enhance accuracy. The present

study points the way forward but clearly further research is needed to investigate the variable effects of stimulated recall by comparing different recall procedures on different aspects of language use.

This study does not give a ringing endorsement to task repetition by itself although it should be noted that the findings are limited by the non-equivalence of the two tasks used in the study. The justification for task repetition as a pedagogic device must lie in whether its effects transfer to a new task (i.e., impacts on development). There was some evidence that the exact repetition in this study may have helped to enhance fluency in a new task but none to show that it contributed to development in complexity or accuracy. The results of this study, then, replicate those of other studies that have reported minimal transfer effects. Perhaps, though, other types of task repetition (i.e., content repetition or procedural repetition) will prove fruitful as indeed Patanasorn (2010) found.

However it is possible that task repetition of any kind will work better when it is combined with an intervention that directs learners' attention to their choice of language (i.e., promotes formulation). Arguably, the stimulated recall in this study did not do this. But conducted in a manner that induces a focus on form – for example as in Fukuta (2016) or Lynch (this volume) – it might do so. Interventions that provide learners with input relevant to the performance of the task or that offer delayed corrective feedback might have a clearer impact on complexity and/or on accuracy (see Sheppard, 2006).

References

Adams, R. (2003). L2 output, reformulation and noticing: Implications for IL development. *Language Teaching Research*, 7(3), 347–376. https://doi.org/10.1191/1362168803lr1270a

Ahmadian, M. J. & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research*, 15(1), 35–59. https://doi.org/10.1177/1362168810383329

Baddeley, A. (2007). *Working memory, thought, and action*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198528012.001.0001

Baleghizadeh, S. & Derakhshesh, A. (2012). The effect of task repetition and noticing on EFL learners' oral output. *International Journal of Instruction*, 5(1), 141–152.

Birjandi, P., & Ahangari, S. (2008). Effects of task repetition on the fluency, complexity and accuracy of Iranian EFL learners' oral discourse. *Asian EFL Journal*, 10(3), 28–52.

Bygate, M. (1996). Effects of task repetition: Appraising the developing language of learners. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 136–146). Oxford: Heinemann.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks, second language learning, teaching and testing* (pp. 23–48). Harlow: Longman.

Bygate, M. (2006). Areas of work that influence L2 speaking instruction. In E. Uso-Juan & A. Martinez-Flor (Eds.), *Current trends in the development and teaching of the four language skills* (pp. 159–186). Berlin: Mouton DeGruyter. https://doi.org/10.1515/9783110197778.3.159

de Jong, N. & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, 61(2), 533–568. https://doi.org/10.1111/j.1467-9922.2010.00620.x

Egi, T. (2007). Interpreting recasts as linguistic evidence: The roles of linguistic target, length, and degree of change. *Studies in Second Language Acquisition*, 29(4), 511–537. https://doi.org/10.1017/S0272263107070416

Egi, T. (2008). Investigating stimulated recall as a cognitive measure: Reactivity and verbal reports in SLA research methodology. *Language Awareness*, 17(3), 212–228. https://doi.org/10.1080/09658410802146859

Ellis, R. (2005). Planning and task-based research: theory and research. In R. Ellis (Ed.). *Planning and task-performance in a second language* (pp. 3–34). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.11.03ell

Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied linguistics*, 30(4), 474–509. https://doi.org/10.1093/applin/amp042

Ellis, R. (2015). *Understanding second language acquisition* (2nd ed.). Oxford: Oxford University Press.

Ellis, R. (2003). *Task-based Language Learning and Teaching*. Oxford: Oxford University Press.

Fukuta, J. (2016). Effects of task repetition on learners' attention orientation in L2 oral production. *Language Teaching Research*, 20(3), 321–340.

Gass, S., & Mackey. A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gass, S., Mackey, A., Alvarez-Torres, M. J., & Fernández-García, M. (1999). The effects of task repetition on linguistic output. *Language Learning*, 49(4), 549–581. https://doi.org/10.1111/0023-8333.00102

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24(2), 95–112. https://doi.org/10.1007/BF02289823

Hant, K. (1966). Recent Measures in syntactic Development. *Elementary English*, 43(7), pp. 732–739

Hawkes, M. L. (2011). Using task repetition to direct learner attention and focus on form. *ELT Journal*, 66(3), 327–336. https://doi.org/10.1093/elt/ccr059

Kim, Y. (2013). Promoting attention to form through task repetition in a Korean EFL context. In K. McDonough & A. Mackey (Eds.), *Second language interaction in diverse educational settings* (pp. 3–24). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.34.04ch1

Kim, Y., & Tracy-Ventura, N. (2013). The role of task repetition in L2 performance development: What needs to be repeated during task-based interaction? *System*, 41(3), 829–840. https://doi.org/10.1016/j.system.2013.08.005

Levelt, W. 1989. *Speaking: From intention to articulation*. Cambridge: Cambridge University Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1–38. https://doi.org/10.1017/S0140525X99001776

Lightbown, P. 2008. Transfer appropriate processing as a model for classroom second language acquisition. In Z. Han (Ed.), *Understanding second language process* (pp. 27–44). Clevedon: Multilingual Matters.

Lindgren, E., & Sullivan, K. P. (2003). Stimulated recall as a trigger for increasing noticing and language awareness in the L2 writing classroom: A case study of two young female writers. *Language Awareness*, 12(3–4), 172–186. https://doi.org/10.1080/09658410308667075

Lindgren, E. & Sullivan, K. (2003). Stimulated recall as a trigger for increasing noticing and language awareness in the L2 writing classroom: A case study of two young female writers. *Language Awareness*, 12, 172–186. https://doi.org/10.1080/09658410308667075

Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, 4(3), 221–250.

Manchón, R. M. (2014). The distinctive nature of task repetition in writing. Implications for theory, research, and pedagogy. *ELIA*, 14, 13–41. https://doi.org/10.12795/elia.2014.i14.02

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139174909

Nation, P., & Newton, J. (2009). *Teaching ESL/EFL listening and speaking*. New York, NY: Routledge.

Patanasorn, C. (2010). *Effects of procedural content and task repetition on accuracy and fluency in an EFL context* (Unpublished PhD dissertation). Northern Arizona University.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.

Robinson, P. (1995). Task complexity and second language narrative discourse. *Language Learning*, 45(1), 99–140. https://doi.org/10.1111/j.1467-1770.1995.tb00964.x

Sheppard, C. (2006). *The effects of instruction directed at the gaps second language learners noticed in their oral production* (Unpublished PhD dissertation). University of Auckland.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language learning*, 49(1), 93–120. https://doi.org/10.1111/1467-9922.00071

Smagorinsky, P. (2001). Rethinking protocol analysis from a cultural perspective. *Annual Review of Applied Linguistics*, 21, 233–245. https://doi.org/10.1017/S0267190501000149

Swain, M. (2006). Verbal protocols: What does it mean for research to use speaking data collection tools. In M. Chalhoub-Deville, C. Chapelle, & P. Duff (Eds.), *Inference and generalizability in applied linguistics: Multiple research perspectives* (pp. 97–114). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.12.07swa

Thai, C., & Boers, F. (2016). Repeating a monologue under increasing time pressure: Effects on fluency, complexity, and accuracy. *TESOL Quarterly*, 50(2), 369–393.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, 24(1), 1–27. https://doi.org/10.1093/applin/24.1.1

CHAPTER 8

Perform, reflect, recycle

Enhancing task repetition in second language speaking classes

Tony Lynch
University of Edinburgh

In this chapter I review my research into the benefits of various forms of internal repetition in speaking tasks created for English for Academic Purposes (EAP) and English for Specific Purposes (ESP) courses at the University of Edinburgh. A common objective of the tasks was that they should create a context for language learners to *recycle* their performance – adapting and modifying their previous output, rather than merely doing the same thing again. I discuss four different speaking tasks and analyze the specific forms of recycling that they involved. After summarizing the evidence that each promoted improved learners' spoken output in later task cycles, I contrast the findings with those from my recent replication study. In that study, I used one of the speaking tasks in a different pedagogic context, and found virtually no change in learners' performances. I will suggest how this difference in uptake might stem from the way in which participants translated task-as-workplan into task-in-process, as well as from other contextual factors. Drawing on this synthesis of two decades of research, I propose a model of task recycling that incorporates the use of reflection activities, feedback, and comparator recordings, through which teachers can help second language learners to recycle and enhance their spoken English output.

Introduction

Teachers tend to think that learners will not welcome being asked to do a task again. Indeed, one specific type of repetition was used as a punishment at my secondary school in the 1960s – namely, having to write out 100 times a sentence related to an alleged misdemeanor, such as "I must not whistle in the school corridors". My approach to persuading learners of English in Edinburgh that repeating a task can be a positive experience has been to create what I call *enhanced repetition*, in which the circumstances of the second round (which I discuss in detail below) are different

DOI 10.1075/tblt.11.08lyn

enough for them not to perceive their second or later performance as merely the same thing again. For the task designer or teacher, the key thing is that the opportunity for learners to repeat a task should bring tangible benefits in the form of improved second language performance, however defined. The same is no less true for the learners, and persuading them that they stand to gain from repeating a task should be easier if the teacher is able to point to firm evidence that repetition has helped other language learners produce better spoken English. It is with learners in mind that I have carried out a number of studies at the University of Edinburgh in search of evidence that the enhanced repetition built into particular speaking tasks did indeed lead to measurable improvements in participants' spoken English. In this chapter I review the findings of my research into four types of oral communication task, each featuring its own form of enhanced repetition. Before discussing those studies, I will briefly set my work against the wider background of research into second language task repetition.

Alternative perspectives in task repetition research

Most research into the benefits of task repetition has prioritised one of two broad aims: *experimental control* or *ecological validity*. Researchers opting for experimental control have tended to set up laboratory-type conditions in which language learners, working alone, perform non-reciprocal speaking tasks to a microphone, having been instructed to imagine they are talking – describing a picture, telling a story, and so on – to an interested listener (Ahmadian & Tavakoli, 2011; Birjandi & Ahangari, 2008; Gass, Mackey, Fernandez, & Alvarez-Torres, 1999). In some cases, such as Bygate (1996, 2001) the learners have performed the focal task direct to the researcher. Paradoxically, in one study the participants were told to "describe the picture, talking as much as possible and *not interacting with the researcher*" (Finardi, 2008, p. 35, my italics). The choice of a non-interactive task format has been made with control in mind: "oral narrative, by virtue of its very monologic nature, induces learners to produce stretches of language that are not influenced by interactional variables" (Ahmadian & Tavakoli, 2011, p. 45). In a limited number of cases, the learners under study have performed tasks interactively in pairs, using a format such as Spot the Difference, where one can expect relatively equal contributions from the two partners (e.g., Gashan & Almohaisen, 2014).

Researchers who have aimed for ecological validity have set up studies based in the classroom rather than the laboratory and have worked with 'intact classes', rather than with groups of individuals assembled for research purposes. In some cases, the focal tasks have been additional to the class's usual work (e.g., Hawkes, 2012; Lynch, 2001) and in others the researchers have investigated tasks that were already part of the class's routine course work (e.g., Lynch & Maclean, 2000, 2001; Nakamura, 2008).

Whichever of these orientations – to experimental control or to ecological validity – has been adopted, most studies have focused on *the linguistic product* of a task in search of empirical evidence of improved performance in a second or later task cycle. One might characterise this as research from the researcher's perspective. Far fewer studies have explored *learners' engagement in the process* of repeating a second language task. Those researchers who have attempted to capture learners' experience of task repetition, through the use of questionnaires and interviews (e.g., Hawkes, 2009; Lynch & Maclean, 2003; Ogilvie & Dunn, 2009; Pang & Skehan, 2014; Pinter, 2005) have found strongly positive perceptions among the learners involved, which suggests that the concerns over possible student boredom and disaffection, originally expressed by Gass et al. (1999), might be misplaced. For example, Hawkes (2009) reported that negative comments on the experience of repeating a speaking task were made by fewer than one in five of the Japanese learners of English in his study. Moreover, since the participants in the study were adolescents, one might have expected them to display a lower boredom threshold than the older learners who are more typical participants in task repetition research. The key to achieving such positive reactions from learners to repeating a task seems to be that they themselves recognised that their first experience of the task helped them put in a better second performance. In the words of one learner, "First time just, just think how to describe the story and ignore the grammar and vocabulary. Second time we think about that grammar and vocabulary and then use the better way to describe the story" (Ogilvie & Dunn, 2009, p. 21) – which simply and succinctly encapsulates the argument developed over many years in the task repetition literature. Pinter (2005) reported similarly positive responses from even younger language learners to tasks incorporating repetition.

A different strand of research, which can be traced back to a sociocultural frame of reference (e.g., Frawley & Lantolf, 1985), emphasises the contribution of the learners to task interaction and their cognitive engagement during repetition of a language task (e.g., Mennim, 2003, 2012). In their Vygotskyan analysis of second language performance, Frawley and Lantolf argued that a first attempt at a second language task (or indeed any task at all) is, for the learner involved, a heuristic exploration of what exactly the task comprises and how to carry it out. Therefore, it makes good sense if second language researchers allow a learner to have at least one dry run before recording their output as 'typical' of that speaker. In Vygotskyan terms, repetition is indispensable for learning; learners must be given the chance to iron out some of the difficulties arising from the unfamiliarity of a task, first time round, in order to be freer (i.e., to have more cognitive processing space free) to express themselves in second and subsequent performances.

Enhanced repetition

In this chapter I will be using the term *enhanced repetition* to refer to the second performance that a learner produces after having had the opportunity to engage in some sort of cognitive activity related to their first run. Such activity can, for example, take the form of feedback (from teacher, peer, or self), or the revision of a transcript of their spoken output (e.g., Li, 2014; Lynch, 2007; Mennim, 2012). It could also involve listening to or watching a comparator recording of a native-speaker or non-native speaker performance of the same task (Lynch & Anderson, 1992; Richards, 1985; Willis, 1996). Current digital technology would make it possible for language teachers to create a library of comparator performances of second language speaking tasks, which learners could access before or after doing a task for the first time (Lynch, 2009a).

The common aim of these various forms of enhancement is that they create the context for a second attempt at a task which does not seem *to the learners* to be simple repetition of what they did before. In a number of research studies, discussed in the next section, I have looked for evidence of both perceived and observed improvements in learners' performance during tasks featuring enhanced repetition.

Benefits of enhanced repetition: Four classroom tasks

Having sketched the theoretical background to task repetition, I now turn to research at the University of Edinburgh into different forms of internal repetition in ESP and EAP speaking tasks created for various courses at the University's Institute for Applied Language Studies (IALS) and its successor, English Language Education. I discuss the findings of research into four different speaking tasks: *Free Talk*, which features repetition with a second group of peers; the *Poster Carousel*, involving repetition with a series of different individual partners; the *Conference Presentation*, which incorporates teacher and peer feedback before repetition; and the *Scenario*, in which learners record, transcribe and revise their own spoken output before the next performance cycle.

Task 1: Free Talk

In 1995 IALS set up an in-session course, the Discussion Group, intended to help international students interact more easily with other students and staff. Its main aim was to give students the opportunity to practise informal speaking and to receive teacher feedback on their spoken output. It would not focus on specifically academic oral skills, which were already addressed in a course in Presentation Skills.

The key question for me as course designer was what to choose as themes for discussion. I assumed that everyone joining a voluntary course would have *something* to talk about and that it was up to me to find a way of giving the learners the chance to talk about whatever that something was. The solution I opted for was on the lines of what would later be known as 'Dogme teaching' (Thornbury, 2000): to prepare no materials, to leave the choice of topics entirely to the students and to make that choice spontaneous, by asking them to say at the start of each lesson what issue interested them that afternoon.

I opted for a group-based discussion format, with learners working in groups of three, and I would leave it up to each trio to decide on the order in which they talked about their topics and how long they spent on each one. These basic design decisions drew on two main strands of research: firstly, the evidence from cognitive psychology (e.g., Craik & Lockhart, 1972; Stevick, 1976) that new language items are best retained in long-term memory when the learner is actively engaged with the learning material and a personal investment in the outcome; and secondly, the argument from SLA studies that, in addition to active involvement in what is being talked about, language learners need "abundant opportunities to control the topic of conversation and self-initiate in class" (Johnson, 1995, p. 85). The task, which I called 'Free Talk', comprised three stages in each 90-minute session:

> Stage 1: Questions and grouping (10 minutes)
>
> Each student wrote down a question to which they wanted an answer, or a problem for which they wanted a solution. The sole criterion was that it should be something they were genuinely interested in. I then grouped the learners in trios, with a mix of topics for each.
>
> Stage 2 – Discussion and monitoring (two cycles of 25 minutes)
>
> Before the task, I asked the learners to monitor how understandable their own speech was for their two partners. The trios then worked in parallel for the next 25 minutes, discussing their questions. I listened in and made a note of language points I thought might be worth commenting on at Stage 3 – not just mistakes, but also expressions that other students might be unfamiliar with. After 25 minutes I got the students to form new trios, to give everyone the chance to try out their question on two new partners, and asked them to keep in mind the points that had caused their first partners problems, and to try to improve on them in round 2.
>
> Stage 3 – Feedback (30 minutes)
>
> I asked the learners themselves to report on things they remembered having had difficulty with in the first round and that went better in the second; I also raised some of the points I had noted myself.

My analysis of recordings made of the interactions in the 1996 Discussion Group pointed to the learning benefits of task repetition in the second round of discussion, which is illustrated in the extract below. It arose from the question *'Which places in*

Scotland should I visit?', asked by a Japanese student, Yuko. In the first round her partners were Isabel (Spanish) and Khalid (Malaysian).

Extract 1.

Yuko: *I have never travelled in Scotland*
Isabel: *hm*
Yuko: *but many people suggest Skye island*
Isabel: *oh I have been there*
Yuko: *oh you have been there* ?
Isabel: *yeah it's really nice it's lovely*
Khalid: *Skye* ?
Isabel: *it's an island* /'aisl ənd/ (corrects herself) *island* /'ail ənd/
Yuko: *island*
Khalid: *oh island*
Yuko: *west ++ northwest of uh + main island*
Isabel: *it's quite nice and you can see ++ seals ? + you know what that is ? ++ seals ? seals*
Yuko: */si:les/* ?
Isabel: (writes on paper) *S-E-A-L-S*
Yuko: *seals* ?
Isabel: *yes*
Yuko: *I've no idea*
Isabel: *they are*
Khalid: *oh an animal* ?
Isabel: *yes*
Khalid: *uh* ?
Isabel: *kind of like this* (starts drawing) *oh + it's not a very good picture but + do you know what I mean* ?
Yuko: *a whale* ?
Khalid: *yeah*
Isabel: *no not like a whale no*
Khalid: *yes + is it like a sea lion ? + or*
Isabel: *yeah kind of I mean*
Khalid: *that + that type*
Isabel: *that type yes ++ yeah they are + outside the water*
Yuko: *is it animal* ?
Khalid: *yeah*
Isabel: *yes but it's not a very good picture*
Yuko: (laughs)
Isabel: *I don't know + how to make + to draw it*
Yuko: *I don't understand ++ this animal can play* ?
Isabel: *no they're + not dolphins*

Khalid: *that' a dolphin* ?
Isabel: *no they're not*
Khalid: *dolphin is bigger than +*
Yuko: *yes I know dolphin*
Isabel: *ah + yes but they're not dolphin + penguins* ?
Khalid: *yeah* ?
Isabel: *no they're not penguins* (laughter)
Yuko: *oh + is it 'seals' in English* ?
Isabel: *yes*
Yuko: *ok + how can we ++ can we see seals at Skye island ? ++ can we see seals* ?
Isabel: *yes it's very nice + I was there last year*
Yuko: *which season did you go there* ?
(etc.)

With the experience of that painstaking first-round negotiation behind her, Yuko moved on to talk with her second-round partners Martin (Swiss) and Iannis (Greek), and was able in the second round to use the lexical item 'seal' successfully and without hesitation when Martin mentioned Skye:

Extract 2.
Yuko: *where do you recommend me to visit in Scotland*?
Iannis: *this is the first time I am in Scotland so I can't…*
Martin: *yeah I have been here in Scotland in a school trip + summer trip ++ the north is a fantastic landscape so + you must go there once*
Yuko: *can you give me + exact advice*?
Martin: *I can recommend Skye + and + also Mull*
Yuko: *Skye and…*?
Martin: *Mull ++ another island +++ and Iona + very beautiful*
Yuko: *I think I must go to Skye ++ you can see many animals ++ like* ***seals*** *+ at Skye island*
Martin: *that's it + and eagles and + sometimes deers + but you must travel by car ++ it's the best way*
(etc.)

Free Talk seems to create relatively natural conditions for learners to practise and improve their spoken English through interactive negotiation of meaning and, in doing so, to co-construct discourse that resembles ordinary conversation. (For more detailed analysis of the interaction between Yuko, Isabel and Khalid, see Lynch, 2009b, pp. 112–113). The task framework gives priority to individual choice of topic and legitimises – and even requires – a type of learner input that in a conventional lesson (with teacher-led input) might be rejected by the teacher as off the point: learner talk

about what currently puzzles, concerns or annoys them. The Discussion Group data show that task repetition can create the opportunity for learners to deploy their linguistic resources more effectively and accurately than in their initial performance.

The obvious question that arises is: if the learners are free to nominate their own topics and are also responsible for such issues as managing time spent on each topic, is Free Talk actually a *task* at all? I think it is. If we take Ellis's definition of a task (Ellis, 2003) as involving

- a primary focus on meaning
- a gap of some kind (information, reasoning or opinion)
- learner choice over the linguistic resources needed to complete the task
- resemblance to real-world language use
- a clearly defined, non-linguistic outcome

I would argue that Free Talk does qualify as a task in terms of the first four criteria, and sometimes the fifth. It requires a primary focus on achieving mutual understanding of the nominated topic, involves a gap of opinion or information (depending on the subject), leaves learners free to deploy their English resources, and leads to exchanges very much like everyday conversation. As to the final criterion, whether a Free Talk discussion leads to a clear non-linguistic outcome varies from topic to topic. To take some examples from my Discussion Group transcripts, 'Where can I buy a second-hand bike in Edinburgh?' and 'Which places should I visit in Scotland?' did produce concrete recommendations. On the other hand, the questions 'Are the British really polite?' and 'What's the point of university?' seemed to reflect an individual's interest in the process of communicating in English, rather than in achieving a defined outcome.

Task 2: Poster Carousel

The Poster Carousel task (Lynch & Maclean, 1994) was originally designed for an *English for Medical Congresses (EMC)* course for health professionals needing to present conference papers in English. The course was 'task-based' in that it involved a series of linked activities over three days that built up to a final-day conference, where research papers were presented, discussed and evaluated. During the course the participants worked through various tasks preparing them for their presentation, for example, describing and discussing data in graphic form. The principal purpose of the Poster Carousel was to provide practice in dealing with audience questions of the sort that might arise in their conference presentation.

The Poster Carousel comprised the following stages:

1. Pairs of participants were given a research article and prepared a poster based on its content.
2. The posters were displayed round a large room. From each pair, one partner stood beside their poster, waiting to receive 'visitors' asking questions. The other partner visited the posters one by one, clockwise, asking questions about each poster. Poster hosts were instructed not to initiate, but to respond to questions. Visitors were allowed limited time (approximately 3 minutes) at each poster.
3. When the visitors arrived back at base, they hosted their poster and their partner went visiting.
4. Once the second round was completed, the participants discussed the merits of the posters, and the teachers provided feedback on general language points.

Repetition in the Carousel is not mere duplication of performance; each successive question-and-answer cycle with a new visitor represents *recycling*, or *re-trial* (Johnson, 1996), where the communicative aim remains the same, but with variations of content and language appropriate to the current visitor's questions.

Having used the Poster Carousel in several *EMC* courses for the Dutch Heart Foundation in the Netherlands, my IALS colleague Joan Maclean and I had informal evidence that it was successful as communicative practice. Our interest in assessing the possible results of recycling more formally led to research (Lynch & Maclean, 2000, 2001) which explored Carousel recordings with two questions in mind:

1. Did learners gain from repetition in the Poster Carousel – and did they *think* they gained?
2. In what ways did they gain from repetition – and in what ways did they *think* they gained?

The participants in our research were 14 oncologists and radiotherapists attending *English for Cancer Conferences (ECC)*, a specialist version of the course which ran in Edinburgh the week before an international conference in the city. They came from six European countries and ranged in English proficiency from around IELTS 4.0 to 7.0.

We collected two types of data. Firstly, we audio-recorded all the interactions between each host and their six visitors; such recording was a routine part of the course. All fourteen sets of six interactions were transcribed for analysis. Our second source of data was a self-report questionnaire, which *ECC* participants completed after the Poster Carousel but before the teachers had commented on their performance. The aim of the questionnaire was to capture their experience of the task as learners.

We analysed the poster hosting performances of five participants, selected to represent the full range of English proficiency within the *ECC* group: 'Alicia' (Spanish; around 4.0 IELTS), 'Susanna' (Hungarian; 4.5); 'Olga' (Serbian; 6.0), 'Carla' (Portuguese; 6.0) and 'Daniela' (German; 7.5). We looked for evidence of varying

effects at different proficiency levels, since language teachers often assume that a task-based approach may work better at higher levels of proficiency, and that tasks should be graded and matched to learners' level (Brown, 1988).

All five participants' recordings demonstrated improvements over the six cycles. Alicia improved the accuracy of her subject-verb structures, aspects of her lexico-grammatical output, and her pronunciation. Susanna showed changes in phonology, syntax and lexis. Olga's performance improved in terms of syntax, lexical selection and phonology. Carla made successful changes in the area of lexis, syntax and fluency of expression. Finally, Daniela self-corrected vocabulary and pronunciation, and also made improvements in the precision and conciseness of explanations.

Were the learners *aware* of the changes and improvements in their successive performances? The questionnaire responses provided some insights into individual perceptions during the Carousel, though they were more limited than we had hoped: although they identified types of change they had made, for example, in grammar or pronunciation, none of the participants gave specific examples of such changes. At the lower end of the range, Alicia was certain that she had neither planned nor noticed any changes during the task, and Susanna told us that she could not remember doing either. However, the recordings showed that both Alicia and Susanna had, in fact, corrected and improved aspects of their English in subsequent performances, even if they were unaware of having done so. Two of the three higher-level learners, Olga and Daniela, reported making conscious changes to their English during the task, and Daniela said she had also noticed making unplanned changes.

If there is an effect of level here, language teachers face the problem that weaker learners may be so focused on making themselves understood that they are not able to notice the changes and improvements they are in fact making to their second language output. To me, that underlines the need for post-task activities that help learners to monitor ways in which their performances have in fact become more accurate, but which they were too preoccupied to notice at the time, in the heat of communicative effort.

The type of interaction created in the Poster Carousel is of particular importance from a pedagogic point of view. Each visitor presents the hosts with a relatively natural opportunity to recycle some of what they have said before, but at the same time the hosts have to tailor their answers to the specific questions asked. The visitor is not simply a listener, but a full conversational partner, initiating topics-for-talk and evaluating the comprehensibility and adequacy of the host's responses.

One other point to stress is that the carousel task was designed to run without teacher intervention. The changes that the *ECC* learners made to their spoken output came about, not in response to external intervention before or during the task, but in pursuit of communication with a series of different visitors. In his study of task repetition with the same partner, Bygate (1996) had speculated that there might be benefits

in getting students to repeat a similar task with a different partner, on the basis that 'different people will do tasks in different ways and a variety of partners could provide valuable learning opportunities' (Bygate, 1996, p. 145). The *ECC* Poster Carousel findings supported that view. What makes the Carousel successful as a communicative task is its particular combination of text input, task structure and learner interaction, which pushes the hosts towards more accurate performance.

We have evidence that the same task can be productive for learners at different levels, including those with very limited English. Teachers who are sceptical about the role of task-based learning at lower levels of proficiency may be encouraged by the fact that even Alicia was not apparently overwhelmed by a task that many native speakers find challenging, namely, responding effectively to probing questions from a fellow-professional at a conference.

Arguably, the learners maintained their level of interest and engagement over the six cycles of the Poster Carousel precisely because they *did not perceive* the task to be repetitious. The recycling built into the task seemed natural and plausible to them, since it derives from a professional genre, poster discussion between host and visitor, of which they had first-hand experience. In addition, it could well be that aspects of the ESP context – e.g., the learners' specialist knowledge, the currency of the poster topics, the additional 'edge' of having to defend the study they had turned into a poster – contributed to their commitment to the task.

Task 3: Conference Presentation

In the Free Talk and Poster Carousel tasks the changes the learners made to their spoken output in a later round of performance arose from learner-to-learner talk in which the teacher did not intervene in any way. I now turn to the third of the four IALS tasks, the Conference Presentation, in which repetition was preceded by teacher intervention and learner reflection.

The evidence for linguistic improvements arising from the Conference Presentation task comes from a different *EMC* course, held in the Netherlands in 2000. There were several reasons for choosing this particular course as our research site. Firstly, feedback on spoken performance was central to the *EMC* rationale and design. Secondly, the feedback procedures combined written teacher records and audio- and video-recordings, so our data collection would not distort the normal course routines. Thirdly, *EMC* catered for a similar learner community to that featured in our Poster Carousel research.

The version of *EMC* that ran in the Netherlands was a 25-hour course over four days. The participants were researchers into cardiovascular health; most were Dutch PhD students, with advanced to native-like proficiency in English. Like *ECC*, the course comprised a series of speaking tasks, culminating in a conference on the final

day. These were preceded, on the first day, by sessions on grammar and pronunciation issues for Dutch users of English, collated from our feedback notes on previous *EMC* courses at the same venue. The points covered in these sessions could be seen as *pre-casts* (Samuda, 2001): Intended to raise awareness, they also represented a practical reference point and baseline, so that feedback could be given more quickly later in the course.

The tasks were designed to increase in difficulty, and the resulting teacher feedback was also designed to increase in complexity. Day 1 featured an audio-recording of problem words. On Day 2, the participants recorded 3-minute oral summaries. On Day 3 group presentations were video-recorded in the morning, and rehearsals (to a teacher) of individual conference presentations were video-recorded in the afternoon. Day 4 was devoted to the final conference, at which each presentation was video-recorded. After the course, the participants were sent a video of their own presentation, with written teacher feedback. Continuous individual feedback records were kept throughout the course. The groups rotated among the three tutors teaching the course, who each added to the feedback record, which was kept by the learner. The learners completed Language Logs when preparing their rehearsal performance and final performance. The question we set out to investigate in our study was: Was there evidence that *EMC* learners (a) become aware of, and (b) master the language problems highlighted in teacher feedback? The data for analysis comprised post-course questionnaires on participants' perceptions of gain and value of feedback; feedback forms and audio-recordings from tutors on sessions from the first three course days, plus post-course feedback from the Conference on Day 4; participants' two language logs; audio–recordings of *Two-Minute Talks* (Day 1), *Reporting back* (Day 2), *Preparing the presentation* (Day 3); and video-recordings of *Rehearsal* (Day 3), and final *Conference* (Day 4).

Analysis of the questionnaire returns showed that pronunciation gains were noted by 21 of the 24 participants. Of these, 14 specified pronunciation of technical terms, and two mentioned phonological changes. Fluency improvements were specified by only two out of 24, though two more mentioned increased confidence. Gains in grammatical accuracy were noted by eight participants – some from feedback and some from the 'pre-cast' session on grammar. Four mentioned improvements in their use of signalling phrases and five others specified gains in the use of stress for emphasising important words, and phrasing. Both these aspects had been introduced in input sessions and then practised in performance.

Dutch-speaking users of English tend to have difficulty with the fortis/lenis distinction, particularly at the end of word followed by one beginning with a voiced consonant. Examples in our *EMC* data included *gap junction*, heard by the tutor as "gab junction", and *Heart Day*, heard as "hard day". The data show that, even over a relatively short period of course time, learners can be made aware of such problems

and can be seen (and heard) working towards greater accuracy. This is illustrated in the extract below, logging the performance of 'Kate', a Dutch researcher working in cardiovascular health promotion. In talking about her work, Kate needed to refer to various types of awareness-raising event, among which were Heart Days and a dance competition for school pupils. Table 1 shows her production of *heart dance awards* over the three days; each tick represents one production of the expression.

Table 1. Production of word-final /t/ in *heart dance* (Kate)

	Produced incorrectly	**Produced incorrectly, after hesitation**	**Produced correctly, after hesitation**	**Produced correctly**
Day 2 *Informal Talk: Tutor feedback*	"I thought you said hard dance awards"			
Day 3 *Presentation: Rehearsal*	√ √ √		√	
Day 4 *Presentation: Performance*			√	√

We can see that Kate became more accurate in that aspect of her pronunciation. After watching the video of her Presentation rehearsal and getting tutor feedback, she was able to get the expression right, first with some hesitation and then without hesitation, in her Presentation on the final day of the course.

There are important differences in the nature of the speaking tasks and feedback in the Poster Carousel and the Conference Presentation. In the Carousel the feedback on performance is implicit, comes from peers, arises in interactive discussion of the poster, and occurs in real time. In the Conference Presentation tasks in *EMC*, feedback is explicit – at levels 10–12 on the scale of graduated feedback (Aljaafreh & Lantolf, 1994) – comes mainly from the tutor, arises primarily from prepared output, and is given after a time-delay to allow the learner to review the recordings.

Most *EMC* participants improved their 'strike-rate' of correct English forms in their Presentation, in areas mentioned by the tutor. The most striking finding was learners' increased awareness of their language use, which several commented on positively in their questionnaires. Throughout the course, the tutors had emphasised that language learning involves both declarative and procedural knowledge, as opposed to the learning of facts more typical of the fields in which the *EMC* participants specialised. A number of questionnaire comments suggested that some participants now recognised that improvements in their English were likely to build over

time, through a complex interaction of rule-learning and adequate opportunity for proceduralisation:

> Questionnaire Extract 1
>
> *"I think if I say something wrong now, at least I realize it, and the pronunciation of a lot of words I am doing better now. But still not everything, but I hope to improve this in the future, because I know now how to pronounce it correctly"*. (Beryl)
>
> Questionnaire Extract 2
>
> *"I have learnt how to do it, now I have to practice!"* (Ann)

Naturally, one cannot claim that all the changes in performance resulted from tutor feedback. Although our analysis focused on the sequential links between points made in written tutor feedback and the accuracy of those items in subsequent performance, there were a variety of other inputs during the *EMC* course: the input lessons themselves; 'collateral' input from feedback given to other participants in the class; and the feedback that some received from peers in collaborative rehearsal.

Our Conference Presentation study went beyond the very short-term focus of Lyster and Ranta (1997), which had examined immediate post-feedback imitation by learners, and of Ohta (1995), which looked at learners' use of recast expressions within the same lesson. It also extended the time-scale of our Poster Carousel research and focused on potential effects during a task series lasting more than three days. The Conference Presentation data suggest that when teachers provide feedback on spoken performance in written form and on an individual basis, and combine it with recordings of the performance, that has the benefit of making feedback more accessible, more easily retrievable later and possibly more effective, especially in a series of linked pedagogic tasks such as those in this *EMC* course. Although it is true that the course featured an unusual degree of individualised feedback, it was by no means unique in doing so. Many language teaching institutions run courses in presentation skills, during which teachers provide some form of written feedback on learners' spoken English, such as the Seminar Evaluation sheet in Anderson, Maclean and Lynch (2004, pp. 200–201). Further research into the effect of combining written feedback and recordings of learner performance should strengthen the argument that learners stand to gain from tasks that allow some form of private rehearsal leading on to teacher feedback and then to repetition in a public presentation, as in the Conference Presentation study.

Task 4: Scenario

The final IALS speaking task involving enhanced repetition is the Scenario (Di Pietro, 1987), a role play designed for pre-sessional EAP courses, which turns on a conflict of interest between a student and a university staff member, such as when a student asks

a tutor to grant an extension of an essay deadline. The basic classroom procedure is in three stages: Preparation, Performance and Debriefing. For the Preparation stage, the teacher divides the class in half and gives each group the role information for one of the two scenario participants (roles A or B). In their separate groups, the learners read their information and discuss how best to achieve their desired outcome and what expressions they might use to that end. For the Performance, the learners from the two groups form A+B pairs and play the scenario simultaneously in parallel, while the teacher monitors and makes notes. For the Debriefing the learners return to their role group and compare the outcomes of the parallel pair conversations, and the teacher provides comments on language issues noted during the performance.

In the original version of the Scenario (Anderson & Lynch, 1992), reflection was limited to the learners' comparison of their success during Debriefing. I subsequently investigated the effects of extending the reflection element before repetition by getting the learners to transcribe and revise their performances (Lynch, 2001, 2007). Transcripts have the potential to provide a platform for what Swain and Lapkin (1995) call the 'reprocessing' of output. They have the advantage of making learners' speech visible, and not merely audible on tape. My hypothesis was that transcribing their own performances would increase the learners' chances of noticing, remembering and producing the reprocessed forms highlighted in process of transcribing. Among the benefits I noted in my first study (Lynch, 2001) was the opportunity for the *re-negotiation of meaning*, as the participants identified language points which had not seemed problematic during the original performance but emerged as unclear at the transcribing stage.

The extract below will give a flavour of the sort of negotiation that self-transcribing gives rise to. Learner M (Greek) and S (Japanese) had played and transcribed a scenario in which the Student (M) has done poorly in a first-term examination and has been called in to talk by the Course Director (S). Here, M and S were reviewing a point in Transcript 1 where M's words had been "I had a lot of work, the dictionaries". (In the extract, underlining highlights the expressions that M and S were discussing or were now suggesting as replacements for what they had in their transcript).

Extract 3.

M: *so I continue + I continue I had to do a lot of work and comma + comma after work*

S: *hm*

M: *for example for example the dictionaries*

S: *oh you said for example no? + did you say for example*?

M: *no now I'm saying now + now*

S: *oh I see*

M: *for example the dictionaries I had to do a lot of work with dictionaries*

S: *hm + oh you say + you want to say + lots of work*
M: *unknown words*
S: *to do with dictionary*
M: *yes yes*
S: *so + how + why don't you say I had to do + I had to do a lot yes I had lots of this ++ lots of work to do + to consult dictionaries*
M: *yes it's ok but I think this phrase is ok had to do a lot of work ++ I had a lot of work to do it's*
S: *yes yes*
M: *the better yes*
S: *yes to do must be break*
M: *yes and the first phrase and the second phrase uh working with or consulting consulting uh the dictionaries*
S: *yes yes + that makes sense*

Several points in the extract are typical of the transcribing data. Firstly, there are no signs of obvious friction between the task partners; they collaborate on the task in hand and seem motivated to produce the best they can. Secondly, the time and trouble they take over revision is striking: All the learner pairs spent more than 45 minutes on what was very detailed work. Thirdly, learners M and S identify language points which had not seemed problematic during the original performance, but which they now realise are unclear – in other words, the revising activity gives them the chance to re-negotiate meanings, in a way that draws their attention to language form in a relatively natural way, and without the direct intervention of the teacher (for the moment).

In a second study (Lynch, 2007) I compared two alternative forms of transcribing. In the first, learners rehearsed and then recorded task performances in pairs, and later transcribed what they said (Transcript 1), before self- and peer-correcting that into a reprocessed version (Transcript 2), which the teacher reformulated into a 'final' version (Transcript 3), for discussion with the speakers. The second variant involved pairs making a recording of their performance, which the teacher took away to listen to, in order to identify and transcribe incorrect language forms. Each pair received a series of transcribed extracts and had to correct them, before checking their changes with the teacher.

The main difference between these procedures lay in the source of the changes to the pairs' output. In the *student-initiated* (SI) variant, the learners themselves are responsible for transcribing their speech verbatim and then for improving it, once they see it in black and white. In *teacher-initiated* (TI) variant, the teacher listens to learners' recordings, selects and transcribes problematic extracts, which are then given to the learners. My study focused on whether SI or TI would be more effective in helping learners reprocess and remember L2 forms from their own performances.

I compared the results of using SI feedback and TI feedback in two intact classes of postgraduate students attending a 10-week IALS pre-sessional English course. Class 1 followed the SI procedure and Class 2 the TI procedure. To compare the possible benefits of the SI and TI feedback, I collected the following data from the two classes: for the SI group I had Transcripts 1–3, and the audio-recordings made in Lesson 1 and Lesson 3; for the TI group I had recordings from Lesson 1, my Lesson 2 feedback extracts, and the recordings from Lesson 3.

In order to assess any longer-term gains, four weeks after the Scenario task I asked the learners from both classes to come to an additional session. At the start of the session I explained they would be making a further recording. I had given them no advance notice of the repetition so that they would have no reason to review their transcripts before the additional session. I recorded and transcribed these final performances, compared them with the earlier recordings, and found that the learners who had followed the SI reflection and revision procedure achieved greater accuracy (64% correct) in the final task repetition on the points they had reprocessed four weeks earlier than those in the TI class (47% correct). In addition to the performance evidence, SI students also reported a higher level of satisfaction than the TI students with the task procedure they had experienced.

The paired SI procedure created greater opportunities for both self- and peer-correction than was required of the TI learners, since the SI partners had to agree first on Transcript 1 and then on their corrections for Transcript 2. It seems likely that self-transcribing and revision in the SI class engaged the learners in an extended version of the monitoring (output searching and checking) behaviour that is thought to promote the automatisation of second language speaking skills (Bygate, 1998).

The fact that the SI class worked on Transcripts 1 and 2 in pairs, not individually, encouraged them to verbalise the process by which they were deciding how to improve their transcribed performance, thus engaging in the 'language related episodes' that can generate new second language knowledge, or consolidate existing partial knowledge (Swain & Lapkin, 1995). Although the data for my study did not include recordings of pairs' discussion about form, my earlier study (Lynch, 2001) had featured such interaction and found evidence of precisely this sort of co-constructed learning. The fact that the SI learners achieved greater accuracy than the TI group suggests that factors other than just repetition were at play. I would argue that one such factor is that self-transcribing of extended discourse involves greater depth of processing (Craik & Lockhart, 1972) than dealing with the teacher's selective transcripts, and hence lays down a stronger memory trace.

Some counter evidence: A second Poster Carousel study

In 2014 I collected data for a new Poster Carousel study, which was intended, broadly, to parallel the earlier *ECC* poster research, but with a number of differences. Firstly, the data for the new study was collected from undergraduates attending a pre-sessional EAP course, rather than the medical professionals on *ECC*. Secondly, the pre-sessional course, *Preparing for Undergraduate Study (PUGS)*, lasted four weeks – four times longer than *ECC* – and allowed time for the carousel task to run twice, in course weeks 2 and 4, which would potentially allow the students to reflect on their performance (by listening to their recordings) between the rounds. Thirdly, while *ECC* focused specifically and explicitly on the improvement of participants' spoken English, the pre-sessional course had a multi-skill focus and was not narrowly targeted on speaking.

The participants in the new study were two classes of Chinese science and engineering students about to enter Year 3 of a four-year degree program at the University of Edinburgh, having completed two years of undergraduate study in China. In preparation for the Edinburgh half of this joint degree, they were required to attend *PUGS*. One element of the course is a 'Research Posters' strand, which occupies 12 of the 80 course hours: four sessions working through introductory materials on preparing an effective poster and locating and converting suitable text into poster form, and then two full mornings (of four sessions each) participating in the Poster Carousel as host and visitor.

In the earlier *ECC* Poster Carousel research Joan Maclean and I had found short-term language improvements over the series of six recordings made over a period of some 20 minutes. This much longer *PUGS* course not only offered scope for more extended poster conversations (of seven to eight minutes) but would also allow comparison of each learner's performances within each round of the Poster Carousel and between the two rounds, held two weeks apart. For the second Poster Carousel the classes were remixed to provide a new set of visitors. In Round 1 the *PUGS* students hosted their poster for members of their own class, while in Round 2 each student hosted members of the other class. My intention was that each poster conversation should require interaction with a visitor absorbing the poster information for the first time and with a fresh eye. The *PUGS* study therefore represented a development of the *ECC* research, by offering the opportunity to explore both short-term and longer-term language improvements.

A total of 12 *PUGS* volunteers agreed to record their conversations as poster host for Rounds 1 and 2 on their mobiles and to send me the audio files for analysis. In the discussion that follows I draw on the performances of three (pseudonymised) students: 'Richard', whose poster was on laser-based lightning conductors; 'Harriet', who had researched a wireless device for recharging batteries; and 'Thomas', hosting

a poster on nanotechnology. Analysis of the recordings revealed some striking differences between the *ECC* and *PUGS* performances. In contrast to the improvements made by the *ECC* participants, there were very few positive changes in the successive *PUGS* recordings. Errors of the sort that in *ECC* tended to be highlighted and corrected through interactive negotiation persisted in the *PUGS* carousels, even when they involved something as basic as incorrect lexical choice, as in the following example from Richard's conversation about lightning conductors.

Extract 4.

Richard: *uh you can see clearly in this poster is about backgrounds, principles and experiments of different devices and different applications about avoiding the* ***thunder*** *+ the* ***thunder***

Visitor: *avoiding* ***lightning*** *+ using laser +*

Richard: *yes about using laser and conducted electricity from the* ***thunder*** *and avoid [??]*

Visitor: *avoiding* ***lightning*** *to the house*

Richard: *uh avoiding* ***lightning*** *+* ***lightning*** *yes using the laser*

Visitor: *using laser yes*

Richard: *the laser has a high + a lot of energy so it can uh makes items into ++ uh + electro and ions + ions + the ions can conduct electricity + so the* ***thunder*** *+ the* ***thunder*** *I mean the electricity will conducted from the sky to earth + just using this way it's very simple*

Visitor: *so where you put this laser? In the top of the house + or somewhere else*?

Richard: *just uh + yes and with a [??] to the sky + not directly + but using a mirrors + mirror*

Visitor: *the* ***lightning*** *uh strike the top of the house not the earth? If the light strictly strikes the earth why you need this*?

Richard: *in fact it's avoiding the* ***thunder*** *not to [??] the* ***thunder*** *not to again use the* ***thunder*** *+ so academic + not in the middle of the thunderstorm*

Visitor: *I'm afraid I don't understand*

Richard: *the* ***thunder*** *+ I know what you mean + you mean the* ***thunder*** *will hit the roof + the roof + and the house*

Visitor: *yes*

Richard: *but we do it just before the* ***thunder***

As we can see, Richard responded in only one speaking turn to his visitor's correction of 'thunder' to 'lightning' and then continued to wrongly use 'thunder', despite his visitor's repeated attempts to put him right. After the episode shown above, Richard persisted with 'thunder' on four further occasions with the same visitor and continued to use it with later visitors.

The *PUGS* learners also maintained quite severe mispronunciations through several task cycles. When hosting her poster on wireless recharging, Harriet repeatedly

produced the first vowel of 'wireless' as rhyming with 'fear', rather than 'fire', and did so in both rounds of the carousel, two weeks apart. Interestingly, none of her visitors queried it, or apparently had any problem understanding it, and some even adopted her mispronunciation when using the word themselves.

Similarly, Thomas was completely consistent throughout the two rounds of the task in mispronouncing 'molecular' as 'mocular' and 'precisely' as 'preciously', as in the extract below:

Extract 5.
Thomas: *nanotechnology is the manipulation of object on atomic* ***mocular*** *and* ***supermocular*** *scale + just like this carbon 60 + it is in a very small scale so we can use it for anything and + the earliest widespread description of nanotechnology was linked to the particular technological goal of* ***preciously*** *manipulating objects and molecules for the fabrication of micro-scale products*

The only improvement I was able to find in the three participants' performances over the two rounds of interactions was in Harriet's recordings and related to clarity of content, rather than accuracy of her spoken English. In explaining the potential medical benefits of wireless charging, Harriet made it increasingly clear for her visitors that she was referring to patients fitted with heart pacemakers:

Extract 6.

Round 1, Visitor 3	*… also there are important applications in medical area because you can charge the device under your skin*
Round 1, Visitor 5	*… and also the most important part it was used in medical area it can charge the device under the skin*
Round 1, Visitor 7	*… and also the most important use is in medical area + it can charge charge the device under the skin ++ if you have any such as heart disease maybe you need a battery to make your heart beat + uh you can you can charge the battery*

One option which was open to the Chinese undergraduates, but had not been to the mixed-nationality *ECC* group, was to resort to a common first language if communication broke down. In the extract below, we can see how Harriet's visitor switched to Chinese as a shortcut and Harriet herself used a phonetic version of 'Faraday' to make herself understood:

Extract 7.
Harriet: *uh have you ever heard about wireless charging? ++*
Visitor: *yeah*
Harriet: *have you ever heard about wireless charging*?
Visitor: *yeah*

Harriet: *uh which aspect*?
Visitor: *uh* ++ [speaks Chinese]
Harriet: *solar energy*?
Visitor: *yes* [laughs]
Harriet: *I thought maybe you have heard about* [speaks Chinese]
Visitor: *yes*
Harriet: *so I'd like to introduce wireless charging in more details* ++ *wireless charging* + *what's your major*?
Visitor: *chemistry*
Harriet: *uh wireless charging is new technology that you can charge your* + *maybe your phones uh without any connection between the phones and the charger* + *and uh this is based on Faraday's law of induction* + *have you ever heard about Faraday's law of induction*?
Visitor: ++ *faya*?
Harriet: *Farady* ++ *fa ra di* [Chinese phonetic form of 'Faraday']
Visitor: *fa ra di yes*

The differences I observed between *ECC* and *PUGS* learners' performances in what was basically the same task might be due to any or all of the contextual features that I noted earlier. Firstly, the *PUGS* students were considerably younger than the *ECC* participants and were halfway through their undergraduate degree, while the individuals attending *ECC* were in their thirties and forties, mostly in university posts and with substantial experience in biomedical research and practice. So, the *PUGS* students brought considerably less specialist academic knowledge to the task than was the case for *ECC* participants. Secondly, again related to the difference in age and general experience, the Chinese undergraduates were unfamiliar with the idea of the research poster, unlike the *ECC* group, who had attended meetings and congresses where they had encountered the poster as a professional genre. Thirdly, *PUGS* is a multi-skill course lasting four weeks and, in its 2014 version, contained strands in academic reading and writing, specialist vocabulary, listening to live lectures, discussion, and language learning strategies. As a result, there was a less sustained pedagogic focus on the skills of speaking than in the four-day *ECC* course, where every course session addressed effective oral communication in English in one way or another. Finally, as noted above, the *PUGS* participants had potential access to a common first language when communication problems arose.

In the light of these differences, it is perhaps not surprising that the undergraduates showed less improvement in spoken performance than we found in *ECC*. The task itself may have been more demanding for *PUGS* students, and therefore required more conscious attention to process and away from language, than it had been for the *ECC* group. However, I believe that an important factor was the way in which the *PUGS* learners – and possibly their teachers – *interpreted the task differently* from the

way intended by the task designers. Differing participant understandings of task processes and goals are at the heart of Breen's (1987) analysis of the life cycle of tasks as task-as-workplan, task-in-process and task-as-outcome, and recently Samuda (2015) has expanded the notion of task-as-workplan to suggest that any learning task evolves over time in its instantiations, as teachers adapt and re-use it to suit the context of specific learning groups. Samuda sets out a minimum of four workplans, shown in Figure 1: the original workplan of the task designer, the teacher's lesson plan, the teacher's on-line plan (made, intentionally or not, as the task unfolds in the classroom) and the retrospective workplan, created after reflecting on what happened in class and possibly including changes for future use.

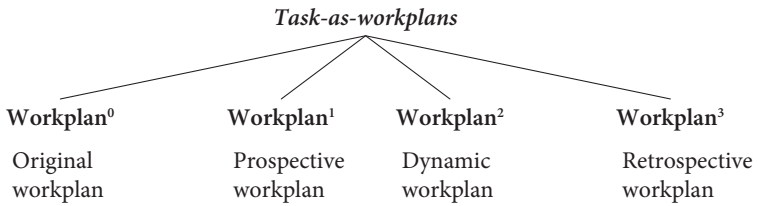

Figure 1. The multidimensional workplan (adapted from Samuda, 2015)

To this teaching perspective, one has to add the differing interpretations of the task that *learners* contribute to the dynamic of the lesson, and the ways in which learners – again, wittingly or unwittingly – can enhance or subvert the intentions of task designer and teacher. Building on earlier work on task instantiation (e.g. Coughlan & Duff, 1994), Gourlay (2005) has explored some of the ways in which control over task interpretation and direction can shift from teacher to learners, in relatively fluid ways "in order to, in the case of the teacher, fulfil their responsibilities and, in the case of the students, negotiate personal agendas" (Gourlay, 2005, p. 421).

As Workplan°, the Poster Carousel as used in the *PUGS* course had undergone several changes over the years since its creation for medical ESP courses. One of those changes was that in *PUGS* the students research and design their posters individually, rather than in pairs. Since the undergraduates come from a range of sub-fields of engineering, as well as chemistry, geophysics and informatics, it had proved difficult in some classes for the teachers to pair up students with sufficiently shared subject backgrounds to work on a poster together. A second change, made in the light of technological progress, is that *PUGS* students create their posters in digital form rather than on paper, as had been the case in *ECC*.

What has not changed in the years since the Poster Carousel was first devised at IALS is its main purpose: to create a communicative context in which the visitors have the opportunity to ask questions and the poster hosts have the opportunity to deal with those questions spontaneously. The instructions in the *PUGS* 2014 students' material read as follows: "As Host, your task is to **respond** to the questions or points

raised by each individual Visitor. Your role is not to initiate; don't give a 'mini-lecture'". The corresponding guidance in the Tutor's Handbook stated: "The Host responds to the questions raised by the Visitor. The point is for them to speak off the cuff. They should not launch into pre-cooked mini-lectures, or read aloud what is on the poster".

However, analysis of the recordings from the *PUGS* Poster Carousel as task-in-process suggests that Richard, Harriet and Thomas interpreted the hosting task as being to give their visitors an 'introduction' (the word they all used) to their chosen topic, rather than to deal with visitors' questions. Unlike the *ECC* participants, they did not allow their visitors time to absorb the poster content and to formulate an opening question. For example, in the extract below we can see how Thomas launched into an exposition as soon as his visitor arrived, before it occurred to him to check what information the visitor actually needed:

Extract 8.

Thomas: *ok so I will begin with the introduction + nanotechnology is the manipulation of object on atomic mocular and supermocular scale + just like this carbon 60 + it is in a very small scale so we can use it for anything and the earliest widespread description of nanotechnology was linked to the particular technological goal of preciously manipulating objects and molecules for the fabrication of micro-scale products often referred to nowadays as mocular nanotechnology +* ***do you want to read this or do you want me to explain this*****?**

Harriet's opening words to a visitor appeared to indicate she was taking his knowledge into account, but then she immediately switched to exposition mode, as a lecturer might:

Extract 9.

Harriet: *My topic is wireless charging and have you ever heard about wireless charging*?

Visitor: *yeah a little bit*

Harriet: *uh wireless charging is a new way to charge a device without connection between device and the charger and uh it is used uh principle of Faraday's law of introduction in- induction induction and uh so it is also called inductive charging ++ well I'd like to introduce Faraday's induction + in more details so you can see these three formulas and uh it is based on these three formulas and there are two coils + uh in the charger and the device and we trans- transmit alternating current AC power to the coil in charger… etc.*

A second difference in approach, adopted by Thomas and Richard, was that they *read aloud from their posters*, again suggesting that they interpreted the task to be one of

presenting the content of their poster, rather than interacting with their visitors, as envisaged in Workplan° by the task designers. The contrasting learner approaches to the Carousel task in the *ECC* and *PUGS* courses may have been related to differences in knowledge status between *PUGS* and *ECC* participants, but also within the *PUGS* group. Firstly, compared with their more experienced ECC counterparts, the *PUGS* students were likely to be less familiar with the academic information they had had to convert into a poster and therefore more dependent on the text of the research article. Secondly, given the differences in specialisms within the undergraduate group, *PUGS* hosts may have assumed that their visitors would not know enough about their topic to understand the poster. In addition, questions from a visitor in difficulty ran the risk of being perceived by the host as a challenge or threat to face. The fact that *PUGS* students resorted to this reading aloud strategy cannot be a simple issue of language proficiency, since the Chinese undergraduates' language level (IELTS 6.5 to 7.5) was in fact higher than most *ECC* participants.

The effect of the host's decision to read aloud from the poster is to reduce or even remove the need for any substantive interaction with the visitor: The host reads aloud and the visitor reads silently. The only reason for the visitor to contribute would be if they wanted an explanation of an unfamiliar expression in the text of the poster – and, again, they may have been discouraged from asking for help if their need for clarification was due to a lack of background knowledge. As a result, the task, as interpreted by these learners, ceases to be a speaking task in any real sense, since what the host is doing is reciting a written text.

As task designer (and author of the *PUGS* syllabus and teaching notes), my response to what I found in the undergraduates' recordings was to look for ways of ensuring that Poster Carousel performances by future *PUGS* participants are more like those in *ECC*. The fact that for the *ECC* participants each new visitor represented a potentially different set of questions, or different variations of the same basic questions, meant that the host's task was to recycle (i.e. adapt for re-use) the information from previous interactions and/or to use new information. It was not a repetition of the same facts in the same order to a different listener, as turned out to be the case for *PUGS* learners.

With this in mind, I made two practical suggestions for change. The first was to make clearer in the *PUGS* teaching notes that the teacher needs to highlight, at every stage of the task cycle, the importance of waiting for the visitor to ask a question: when working through the introductory materials, when opening the Carousel session and when providing feedback after the first round. The second was to recommend that the Carousel session, shown as *Poster Presentation* on the course timetable, should be re-named *Poster Questions and Answers*. These two changes were intended to help remind students (and teachers) that the Carousel is a vehicle for practice in handling visitors' questions – in other words, that the poster should be a springboard

for interaction and negotiation. The recommendations were acted on by the English Language Education staff responsible for the 2016 *PUGS* course, and research is currently under way to see whether the changes I proposed can be shown to have had any positive influence on the interactive balance between poster hosts and visitors.

A model for maximising the benefits of enhanced repetition

Four of these five Edinburgh studies have confirmed, with intact classes, the findings of other research into speaking task repetition under experimental or quasi-experimental conditions, which had reported increased accuracy, complexity and fluency of learners' English on subsequent performance. My research also suggests that the various forms of enhancement built into the tasks were successful in creating a context in which learners did not perceive the second round of performance as mere 'verbatim repetition' (Ahmadian, 2012, p. 380). This raises the issue of how best to help teachers make practical use of such findings in the way they set up and monitor and provide feedback on learning tasks, and I offer Figure 2 as a framework for representing the pedagogic options explored in the Edinburgh task repetition research.

The three main phases in the model are *Input*, where the teacher provides the learners the information and/or texts on which the task is based, *Performance* itself, and *Repetition*, which could be a single second performance or several further performances. After receiving input and before performing the task, learners have three options, which are not mutually exclusive: to plan in general how to approach the task, to rehearse in some detail what they intend to say, and listen to or watch a comparator performance by native or non-native speakers.

We can see that the task in the Edinburgh series with the least explicit form of enhancement is Free Talk, where the teacher asks the learners to monitor the comprehensibility of their contributions to the first task cycle (looking for signs that their two partners are having difficulty understanding them), to see whether they can identify the source of the difficulty (e.g., a lexical gap such as Yuko's 'seals'), and to find a satisfactory solution to the problem, so that in the second round of the task they may be able to make a conscious effort to improve on those points as they interact with new partners.

The Poster Carousel engages learners in a substantial period of Planning (involving reading one or more academic texts, summarizing the content and converting it to graphic form in a poster), followed by Performance as poster hosts with their first visitor and then by Repetition with a series of further individual visitors. At the next stage the teacher provides plenary feedback on points they have noted during the carousel. In the task procedure used in the *PUGS* course, the learners then have the opportunity over the next two weeks to listen to their own recordings of their Round 1 hostings (and to look up, or check with the teacher, language points they

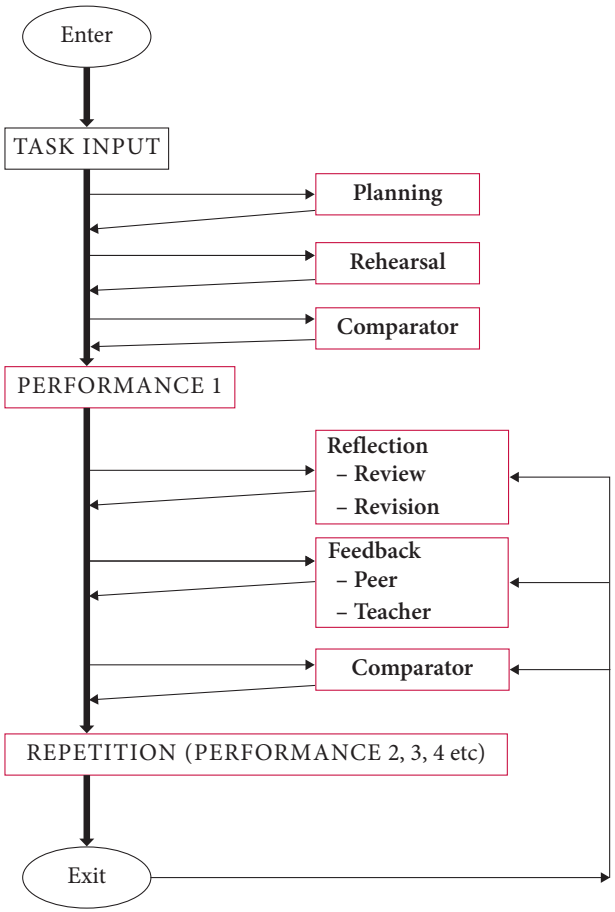

Figure 2. A framework for enhanced repetition in language learning tasks

have doubts about) before they participate in additional Repetition in Round 2 with a further set of visitors.

In the Scenario procedure set out in Anderson, Maclean & Lynch (2004), Input takes the form of role-play information given separately to the two A and B halves of the class. At the next stage, Planning, the A and B learner groups discuss possible tactics for resolving the problem to their advantage and rehearse what they think will be useful linguistic expressions for the coming conversation. Again, there is the opportunity for them to use a comparator recording at this point. For the Performance stage the learners play the scenario in A+B pairs and then return to their own role group to compare the process and outcome of their conversation with those of their fellow group members and receive peer feedback. Here the teacher has the option of playing a comparator performance, allowing the learners to evaluate their own conversation alongside that of another pair. They then move on to Repetition, working

with a different A or B partner, and record their performance. That recording then forms the basis for Revision, in the form of the transcribing activities that I described earlier, leading on to Feedback from the teacher (i.e. comments on their revised transcript) and to further Repetition, initially with their transcript partner and then with other peers.

In the Conference Presentation the task Input varies: It may be the learner's own research publications or other academic texts. Planning of the content and form of the talk is followed by private rehearsal and some *EMC* learners choose to rehearse the talk to a peer as well. Performance involves delivering the talk to the teacher, who records the performance and provides detailed written feedback on points of language (and delivery). The learner then watches the recording of their talk and follows the teacher's notes. Learners can also ask peers to watch their performance and provide further feedback. Repetition in this case takes the form of public presentation of the talk at the end-of-course conference, followed by audience questions and answers (representing an additional type of peer feedback). Finally, in the Dutch version of the *EMC* course, the teachers provided delayed written feedback on a recording of the public talk, which they returned to the learners a week after the course.

Figure 2 is intended to highlight alternative options for pedagogic paths through a speaking task. It shows that repetition can be embedded within different sets of activities and does not have to be limited to doing the same thing again immediately. The timing of the use of comparators, in particular, is meant to be flexible: Recordings can be used either as part of Input, or after Performance, or after Repetition, and – given the appropriate facilities – can be used by learners working in self-access mode or by teachers in class. (For further discussion, see Lynch, 2009a).

Conclusion

I began this chapter by noting the concerns expressed by SLA researchers (mainly) about the risk of learner boredom with task repetition. The Edinburgh studies I have reviewed here, tapping into learners' perceptions of the experience of repeating second language tasks, suggest that such concerns may have been somewhat excessive. It is perhaps worth noting in passing that, although the English lexical network of 'repeat' ('repetition', 'repetitive', 'repetitious', 'repeatedly', etc.) tends to be negatively associated with *unnecessary* repetition, other European languages use cognates of 'repetition' neutrally or positively for 'rehearsal'. Taking a wider perspective, both Japanese and Chinese have separate words for 'repetition', with negative and positive associations. It could therefore be that SLA researchers' concerns over classroom repetition have been rather anglocentric. Second language users of English may be more tolerant of the notion and practice of repetition, as the available research of learners'

perceptions now suggests. Advocates of the sociocultural approach have long argued that repetition and recycling are an essential part of the learning process, and one does not need to subscribe fully to a socio-cultural view of language learning to accept that giving learners a second chance at a communication task is likely to bring about a better second performance.

I believe the key to success, in terms of both process and outcome, of the repetition tasks in four of the five Edinburgh research projects was that the participants recognised (and took) the chance to *recycle* their output, rather than simply repeat it, with sufficient freshness about the second and later task cycles for them to feel a new challenge with a different partner. The exception was the recent study of the Poster Carousel in the undergraduate pre-sessional EAP course, and I have suggested possible reasons for the lack of obvious improvement in those learners' performances. If fears do remain in language teachers' minds over learners' attitudes to task repetition, these Edinburgh findings may help persuade them that a second attempt represents a positive opportunity for language learners to improve on a first performance, particularly when combined with appropriate reflection, feedback and revision.

References

Ahmadian, M. (2012). Task repetition in ELT. *ELT Journal*, 66(4), 380–382. https://doi.org/10.1093/elt/ccs020

Ahmadian, M., & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production *Language Teaching Research*, 15, 35–59. https://doi.org/10.1177/1362168810383329

Aljaafreh, A., & Lantolf, J. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *Modern Language Journal*, 78, 465–483. https://doi.org/10.1111/j.1540-4781.1994.tb02064.x

Anderson, K., Maclean, J., & Lynch, T. (2004). *Study speaking* (2nd ed.). Cambridge: Cambridge University Press.

Birjandi, P., & Ahangari, S. (2008) Effects of task repetition on the fluency, complexity and accuracy of Iranian EFL learners' oral discourse. *Asian EFL Journal Quarterly*, 10(3), 28–52.

Brown, R. (1988). Classroom pedagogics – A syllabus for the interactive stage? *The Teacher Trainer*, 2(3), 8–9.

Bygate, M. (1996). Effect of task repetition: Appraising the development of second language learners. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 136–46). Oxford: Heinemann.

Bygate, M. (1998). Theoretical perspectives on speaking. *Annual Review of Applied Linguistics*, 18, 3–19. https://doi.org/10.1017/S0267190500003469

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 23–48). Harlow: Longman.

Bygate, M., & Samuda, V. (2005). Integrative planning through the use of task repetition. In R. Ellis (Ed.), *Planning and task performance in second language* (pp. 37–74). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.11.05byg

Coughlan, P., & Duff, P. (1994). Same task, different activities: Analysis of an SLA task from an activity theory perspective. In J. Lantolf & G. Appel (Eds.), *Vygotskyan approaches to second language research* (pp. 173–193). Norwood, NJ: Ablex.

Craik, F., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Di Pietro, R. (1987). *Strategic interaction: Learning language through scenarios*. Cambridge: Cambridge University Press.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Finardi, K. (2008). Effects of task repetition on L2 oral performance. *Trabalhos de Linguistica Aplicada*, 47(1), 31–43. https://doi.org/10.1590/S0103-18132008000100003

Frawley, W., & Lantolf, J. (1985) Second language discourse: A Vygotskyan perspective. *Applied Linguistics*, 6(1), 19–44. https://doi.org/10.1093/applin/6.1.19

Gashan, A., & Almohaisen, F. (2014). The effect of task repetition on the fluency and accuracy of Saudi female learners' oral performance. *Advances in Language and Literacy Studies*, 5(3), 36–41. https://doi.org/10.7575/aiac.alls.v.5n.3p.36

Gass, S., Mackey, A., Fernandez, M., & Alvarez-Torres, M. (1999). The effects of task repetition on linguistic output. *Language Learning*, 49, 549–80. https://doi.org/10.1111/0023-8333.00102

Gourlay, L. (2005). Directions and indirect action: Learner adaptation of a classroom task. *ELT Journal*, 59(3), 209–216. https://doi.org/10.1093/elt/cci040

Hawkes, M. (2009). Effects of task repetition on learner motivation. In A. Stoke (Ed.), *JALT2009 conference proceedings* (pp. 456–465). Tokyo: JALT.

Hawkes, M. (2012). Using task repetition to direct learner attention and focus on form. *ELT Journal*, 66(3), 327–336. https://doi.org/10.1093/elt/ccr059

Johnson, K. (1995). *Understanding communication in the second language classroom*. Cambridge: Cambridge University Press.

Li, Q. (2014). Get it right in the end: The effects of post-task transcribing on learners' oral performance. In P. Skehan (Ed.) *Processing perspectives on task performance* (pp. 129–154). Amsterdam: John Benjamins.

Lynch, T. (2001). Seeing what they meant: Transcribing as a route to noticing. *ELT Journal*, 55(2), 124–132. https://doi.org/10.1093/elt/55.2.124

Lynch, T. (2007). Learning from the transcripts of an oral communication task. *ELT Journal*, 61(4), 311–320. https://doi.org/10.1093/elt/ccm050

Lynch, T. (2009a) Responding to learners' perceptions of feedback: The use of comparators in second language speaking courses. *Innovation in Language Learning and Teaching*, 3(1), 1–13.

Lynch, T. (2009b) *Teaching second language listening*. Oxford: Oxford University Press.

Lynch, T., & Anderson, K. (1992). *Study speaking*. Cambridge: Cambridge University Press.

Lynch, T., & Maclean, J. (1994). Poster carousel. In K. Bailey & L. Savage (Eds.), *New ways of teaching speaking* (pp. 278–279). Washington DC: TESOL.

Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, 4(3), 221–250.

Lynch, T., & Maclean, J. (2001). 'A case of exercising': Effects of immediate task repetition on learners' performance. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and research* (pp. 141–162). Harlow: Longman.

Lynch, T., & Maclean, J. (2003). Effects of feedback on performance: A study of advanced learners on an ESP course. *Edinburgh Working Papers in Applied Linguistics*, 12, 19–44.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, 19, 37–66. https://doi.org/10.1017/S0272263197001034

Mennim, P. (2003). Rehearsed oral output and reactive focus on form. *ELT Journal*, 57(2), 130–138. https://doi.org/10.1093/elt/57.2.130

Mennim, P. (2012). Learner negotiation of L2 form in transcription exercises. *ELT Journal*, 66(1), 52–61. https://doi.org/10.1093/elt/ccr018

Nakamura, E. (2008). Effects of task repetition on poster carousel. In K. Bradford-Watts, T. Muller, & M. Swanson (Eds.), *JALT2007 conference proceedings* (pp. 188–199). Tokyo: JALT.

Ogilvie, G., & Dunn, W. (2009, September). Task repetition as a pedagogical strategy: Interactional effects and affective responses. Paper presented at the Third International Conference on Task-Based Language Teaching, Lancaster, England.

Ohta, A. (1995). Applying sociocultural theory to an analysis of learner discourse: Learner-learner collaborative interaction in the zone of proximal development. *Issues in Applied Linguistics*, 6(2), 93–121.

Pang, F., & Skehan, P. (2014) Self-reported planning behaviour and second language performance in narrative retelling. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 95–128). Amsterdam: John Benjamins. https://doi.org/10.1075/tblt.5.04pan

Pinter, A. (2005) Task repetition with 10-year-old children. In J. Willis & C. Edwards (Eds.), *Teachers exploring tasks in English language teaching* (pp. 113–126). Houndmills: Palgrave Macmillan.

Richards, J. (1985) Conversational competence through role-play activities. *RELC Journal*, 16(1), 82–100. https://doi.org/10.1177/003368828501600107

Samuda, V. (2001). Guiding relationships between form and meaning during task performance: The role of the teacher. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and research* (pp. 119–140). Harlow: Longman.

Samuda, V. (2015). Tasks, design and the architecture of pedagogic spaces. In M. Bygate (Ed.) *Domains and directions in the development of TBLT*: *A decade of plenaries from the international conference* (pp. 271–302). Amsterdam: John Benjamins. https://doi.org/10.1075/tblt.8.10sam

Stevick, E. (1976). *Memory, meaning and method*. Rowley, MA: Newbury House.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, 16(4), 371–391. https://doi.org/10.1093/applin/16.3.371

Thornbury, S. (2000). A dogma for EFL. *IATEFL Issues*, 153(2).

Willis, J. (1996). *A framework for task-based learning*. London: Longman.

CHAPTER 9

Second language learning through repeated engagement in a poster presentation task

Emi Kobayashi and Masaki Kobayashi
Kyoai Gakuen University / Kanda University of International Studies

This chapter reports on an ethnographic multiple-case study of EFL students' language learning through repeated engagement in a poster presentation task. The study draws upon sociocultural and ecological perspectives to highlight the interpersonal processes that contribute to L2 students' learning through a poster carousel (Lynch & Maclean, 2000, 2001) that involved three presentations by the same group on the same topic, each followed by a question-answer session. Participants were Japanese undergraduate students, and data were collected primarily through classroom observations, audio-recordings of task-related interactions, semi-structured interviews, and collection of relevant documents. Recorded discourse was analyzed mainly using an ethnography of communication approach (Duff, 2002) in order to trace the interactional trajectories that led to changes in students' performances. The analysis showed that students benefited from their repeated experiences with the task in a variety of ways. Some students acted upon the feedback from their audience to refine their subsequent task performances. Other students used questions from their previous audience to invite contributions from the current audience. Importantly, not all students were willing to change their second and third performances because their main concern was to avoid risk-taking. The chapter concludes by discussing the implications of these findings for research and pedagogy.

Introduction

In recent years, an increasing number of second language (L2) studies have examined the effects of task repetition on learners' L2 production, reporting on meaningful changes (or lack thereof) in linguistic accuracy, complexity, and oral fluency (e.g., Bygate, 2001; Fukuta, 2016; Gass, Mackey, Alvarez-Torres, & Fernández-García, 1999), the three dimensions of L2 production that are claimed to compete for speaker attention during task performance (Skehan, 1998). These studies have contributed to a greater understanding of the relationship between the information processing

DOI 10.1075/tblt.11.09kob

involved in the process of repeating L2 tasks and resultant changes in task performance. However, only a few studies have examined task repetition as it occurs in its natural context (e.g., Hawkes, 2012; Kim & Tracy-Ventura, 2013; Lynch, this volume; Lynch & Maclean, 2000, 2001; Nitta & Baba, this volume; Shintani, 2012, this volume), and not all of these have involved students sharing the same first language (which excludes Kim & Tracy-Ventura, 2013 and Lynch's studies).

The importance of context has long been recognized in the literature. For example, Mercer (1992) stressed the situatedness of learning arguing that "the study of learning, especially in educational settings, must treat context and culture as part of what is being studied, not variables to be partialled out" (p. 33). More recently, Samuda and Bygate (2008) have called for L2 research on tasks contextualized in educational programs, saying that "action on tasks is heavily influenced by the context, since it is the context that provides the primary grounds for the participants' interpretations of what they are doing and why they are doing it" (p. 258). Thus studies conducted in artificial contexts set up for the sake of research may suffer from a lack of ecological validity as they are likely to leave out important contextual and learner contributions to task engagement and performance, including student motives and histories, interlocutor relations, and teacher roles and expectations (see also Bygate & Samuda, 2009; Coughlan & Duff, 1994; Duff, 2007a, for further discussion). This chapter therefore takes an ethnographic approach to explore EFL students' learning through repeated engagement in a particular type of classroom task, namely poster presentations, in their classroom.

As Skehan (1998) has suggested, task repetition can be regarded as comparable to strategic planning that allows learners to consider the content and language of their speech in advance. It is indeed referred to as "integrative planning" (Bygate & Samuda, 2005) in the literature. According to Bygate and Samuda (2005),

> An initial encounter with a task can be seen as creating a holistic representation of the task, along with the experience of handling it in real time. This representation and the accompanying experience can be stored, creating a kind of plan which can be drawn on on a second occasion, enabling the learner to integrate a broader range of their resources into their performance. (p. 38)

Thus, task repetition may allow learners to "perform more adequately their competence as it were" (Bygate & Samuda, 2005, p. 38). While this view focuses on the language use and information processing of individual learners involved in task repetition, this chapter draws upon ecological and sociocultural perspectives (e.g., Coughlan & Duff, 1994; Duff & Kobayashi, 2010; Mohan & Marshall-Smith, 1992; Swain, 2006; van Lier, 2004) to highlight the interpersonal processes involved in repeating a task in pairs or groups that may enable L2 students to produce language that they have not yet fully mastered – or in Swain's (2000) words, to outstrip their

competence – and contribute to their learning across task-related events. In particular, the study examines not only students' actual task performances, but also the collaborative interactions in which they engage with their audience and partners during and between tasks to co-construct meaning in and knowledge about their L2. For the purposes of the study, the task consists of students' group presentations and the ensuing question-answer session.

Previous studies on task repetition

Because of their focus on the interpersonal processes involved in repeating the same task, five previous studies are of particular relevance. Coughlan and Duff's (1994) study showed how the same L2 speaker produced different types of discourse at different times, demonstrating that individuals as active agents can interpret tasks and perform them in ways that are different from those expected by the researcher or teacher. This study also revealed how one of the researchers and her ESL participant co-constructed their task-based interaction dynamically, with each contributing their experience and expertise relevant to the interaction as the interaction unfolded.

Likewise, Pinter (2007) reported a change in her participants' interpretations of a task. Her research examined the interaction of two Hungarian boys doing several versions of a spot-the-differences task in English. The analysis of their interactions together with their reflective comments indicated that these friends learned to attend and respond to each other's contributions over multiple trials. Shintani (2012, see also this volume) also examined task repetition with a group of young Japanese children with no previous knowledge of English doing an input-based task. Her analyses showed that the participants gained greater ease in performing the task and improved their comprehension over nine lessons. Also, the analyses suggested that the teacher adjusted the nature of assistance as needed and that although the task did not require L2 production, the participants spontaneously used a wider range of L2 social speech over time while producing fewer L1 utterances. These findings have implications for the role of teacher scaffolding and of L1 in task repetition.

Perhaps of the greatest relevance to this chapter is Lynch and Maclean's (2001) study that investigated how European students in a medical ESP program learned English through their repeated participation in poster sessions, or what they called "poster carousels" (p. 142; see also Lynch, this volume). The students worked in pairs for one hour to read assigned research articles and make a poster. This preparation phase was followed by the major task phase where one student from each pair stood beside their poster to answer questions from visitors while the other student visited other pairs' posters. This task forced the presenter to answer their peers' questions several times, thus providing a relatively natural context for task repetition.

The researchers note that the carousel differs from monologue tasks used in previous studies on task repetition in that it requires the presenter and audience members to interact with one another, thus making the roles of speaker and listener more fluid. Their analysis indicated that irrespective of level and language used participants produced more accurate language as they became familiar with their message content and that they were also able to use their peers as sources of new or more accurate L2 expressions.

Although not strictly a study of task repetition, Kobayashi's (2016) ethnographic case study followed a Japanese ESL student's longitudinal trajectory across three different presentations given as major assignments of her content courses at a Canadian university – what might be seen as a variety of 'procedural repetition' (see Kim et al., this volume). The analysis showed that she constructed and reconstructed her understanding of what it meant to do an oral presentation (OP) and performed her tasks differently as she became socialized into the socioculturally valued practices of her classroom community over time. Given the paucity of task repetition research with a group of same-L1 students, the present study aimed to add to the limited knowledge in this area by detailing Japanese students' learning through three performances of an OP task in a naturally occurring EFL class at a Japanese university.

Theoretical framework

This inquiry was guided by sociocultural and ecological perspectives (Duff, 2007b; Lantolf & Thorne, 2006; van Lier, 2004). According to Vygotsky (1978), learning and cognitive change take place in the zone of proximal development (ZPD), which can be defined as "the difference between the level of development already obtained and the cognitive functions comprising the proximal next stage that may be visible through participation in collaborative activity" (Lantolf & Thorne, 2006, p. 20). Vygotsky emphasizes the socioculturally mediated nature of learning and the importance of language as a major tool in this process. In other words, it is by means of language that individuals participate in various activities and develop socially valued knowledge and skills (Lantolf & Thorne, 2006). As suggested by Wells (1999), other cultural resources that mediate human actions and interactions include values, beliefs, and attitudes regarding what are socially, culturally important activities to engage in, and knowledge and understanding of the practices of these activities. Moreover, as Wink and Putney (2002) argue, "[w]henever we engage in any conversation, no matter how brief, we begin to build historicity" (p. 135). From this perspective, task repetition can be seen as multiple opportunities to create a shared understanding through language-mediated interactions.

Likewise, the theory of language socialization suggests that L2 learning is an interactional process whereby novices are apprenticed into what Gee (1996) refers to as "Discourses" with a capital D or "ways of behaving, interacting, valuing, thinking, believing, speaking, often reading and writing that are accepted as instantiations of particular roles (or 'types of people') by specific *groups of people*" (Gee, 1996, emphasis original, p. viii). In this view, it is through their repeated exposure to and participation in language-mediated interactions with more experienced members that novices learn the language and culture of their target community (Ochs, 1988). As Wortham and Reyes (2015) put it:

> Learning involves increasingly competent participation in social activities across events, with exposure and practice in one event facilitating participation in subsequent events. Few human processes take place exclusively within single events2. (p. 17; see also van Lier, 1988)

As such, language socialization can provide a useful theoretical lens through which to examine students' learning through task repetition.

In addition, the present study was informed by van Lier's (2000, 2004) ecological-semiotic perspective, which sees L2 learners as situated in a complex web of relations that constitutes the ecology of language learning and teaching. Central to this perspective is Gibson's (1979) notion of *affordance* (see also Nitta & Baba this volume), which is defined by van Lier (2000) as "a property of the environment that is relevant – for good or ill – to an active, perceiving organism in that environment" (p. 252). Van Lier (2004) argues that for L2 learning to occur, individuals "must be engaged in activity and have information around that is available to be picked up and used" (p. 97), suggesting that access and engagement are essential. What this indicates for the present study is that the presenters' interactions with their audience and group members (which form part of the data) may become linguistic affordances.

Furthermore, we draw on four theoretical concepts developed in different but closely related areas of research. The first concept is that of *languaging* (Swain, 2006). According to Swain, this Vygotskian notion refers to "the process of making meaning and shaping knowledge and experience through language" (p. 98). This can be mediated either through collaborative dialogue among members working in a group or private speech. In either case, languaging enables observers to see how learners engage with language and reach a better understanding of initially challenging material.

Second, we adopt Coughlan and Duff's (1994) distinction between the terms *task* and *activity*, which parallels Breen's (1987) distinction between *task-as-plan* and *task-in-process*. Informed by Activity Theory, Coughlan and Duff suggest that the former term refers to a "behavioral blueprint" (p. 175), whereas the latter refers to a resultant experience that task performers construct on the basis of the perceived goals and purposes of the task as well as of their own histories and needs.

The third concept *appropriation* derives from Bakhtin's dialogism (1981):

> The word in language is half someone else's. It becomes 'one's own' only when the speaker populates it with his own intention, his own accent, when he appropriates the word, adapting it to his own semantic and expressive intention. (pp. 293–294)

In other words, appropriation is a process of taking others' words and taking ownership of them to fulfill one's own communicative intentions and purposes.

Fourthly, this research draws on Goffman's (1959) theatrical metaphor, which distinguishes front stage from back stage. These two concepts help us see how people change their behavior depending on the audience involved in the speech event. Actors performing formally in front stage attempt to present idealized versions of themselves as their behavior and performance are open to evaluation by their audience. In contrast, back stage the same actors may reveal sides of themselves that the audience may find unofficial or private. Because backstage areas are "typically out of bounds to members of the audience" (p. 128), actors may step out of character and express themselves more freely. Kobayashi (2004) reported that Japanese ESL students used L1 backstage talk before and during their OPs for different reasons. For example, a male student gave advice to his partner on her volume of speech ("Speak up") and behavior ("You should look up"). This, as Kobayashi claimed, can be considered as a type of "backstage" talk (Goffman, 1959) in that although uttered on stage, it was addressed only to another member of the presenting group, as indicated by the speaker's use of whispering (see also Kobayashi, 2016). It is also a typical kind of "rehearsal" talk whether in theatre or public speaking. Batstone and Philip (2013) make a similar distinction between public and private space, saying that interaction in the former space is "intended to be heard by everyone in the class (in the case of teacher-led discourse) or by everyone in the group (in the case of group-work)", whereas the interaction in the latter space is "intended only for a sub-group" (p. 118). Thus, it is worth trying to apply the following parallel: Frontstage talk takes place in public space whereas backstage talk relates to private space.

Informed by these perspectives, we view tasks in terms of how they are situated in a particular classroom context. Given the Vygotskian notions of co-construction and historicity (Wink & Putney, 2002 see also Mercer & Littleton, 2013), it follows that even when the same task is repeated, each performance generates a unique activity that provides different learners with different linguistic affordances. Moreover, task repetition is deemed to entail students not only having multiple opportunities to language about the same material and appropriate the voices of their interlocutors, but also learning socially and culturally valued ways of doing tasks. As such, it is deemed valuable to examine not just the *microprocesses* of students' L2 discourse within each task performance, but also the *macroprocesses* that can mediate their learning through

task repetition (Mohan & Marshall Smith, 1992), including teacher explanation and scaffolding and behind-the-scenes peer collaboration (Duff & Kobayashi, 2010).

Empirical study

The purpose of the study

Drawing primarily upon the sociocultural and ecological perspectives outlined above, this multiple-case study examines L2 students' learning through a poster carousel that involved three OPs and question-and-answer sessions. More specifically, the study was guided by the following research questions:

1. What linguistic affordances does repeated performance of an OP task provide for Japanese undergraduate students in an EFL classroom?
2. How do they respond to these affordances to improve their task performance?

Context of the study and participants

This chapter draws on a subset of data which was collected for a larger longitudinal ethnographic project (Kobayashi, 2007) that involved two classes of undergraduate students taking a year-long Freshman English (FE) course (April 2005- January 2006) at a private university in Japan. This is a required course for all first-year students majoring in English. It was designed to help students improve their ability to communicate in spoken and written English and become more aware of their own ways of learning by providing them with opportunities to participate in a variety of practical and reflective activities. Although students' use of L1 was tolerated or even encouraged in some cases (e.g., when discussing complex ideas or grammar) by the instructor, the major medium of instruction was English. Classes met once a week for 90 minutes. While all students enrolled in the classes participated in the larger study, this chapter focuses on one of the two classes of 14 students (mean TOEIC score: 610) and presents the experiences of four (one male and three females) and their partners as information-rich cases (Duff, 2008) of learning through repeated task performances in this classroom community. Each focal student worked with a peer or two to do the presentation task. Table 1 presents a brief overview of these students.

Table 1. Focal students' profiles

	Ryo	Nagi	Ako	Rena
TOEIC	**590**	**685**	**650**	**620**
English language learning experience	– Enjoyed taking part in English speech contests and won many of them	– (in her high school days) Performed a short play in English at a public event	– Had some negative experiences using English in lessons with native teachers of English at a private language school	– Struggled with English in junior high school – Felt a sense of accomplishment when she got higher scores – Not confident about speaking English
Experience abroad	– Spent four years at primary school in the USA	– Two-week homestay in the USA	– None	– None – Looking forward to participating in a year-long study abroad program
Future goals	– To become an English language teacher	– To become an English language teacher or a tour guide	– To become an English language teacher like her own teacher	– Not decided yet; Wanted to find out through her undergraduate studies what she could do with English
Task partner(s)	– Moe – Shiho	– Jun	– Hina – Yui	– Mao

The focal task

Below we focus on the first of the two group OP tasks that students in FE undertook over the academic year. This Semester 1 task involved a series of three presentations of the same poster to different audiences. More specifically, students were required to form groups of two or three to read newspaper articles on a current or controversial topic, create a poster out of class, and give a 15-minute presentation followed by a 5-minute question-and-answer session (see Table 2). Unlike the poster carousel used in Lynch and Maclean's (2001) research, the focal task reported in this chapter required both (or, in the case of groups of three, all) of the students to stay beside their poster and give a joint talk. Most of the preparation, including reading and discussing an article, making a poster, and preparing scripts, took place outside the classroom over several days. The group presentation task itself involved presenting a summary of their articles and discussing their positions three times (three iterations or rounds),

as well as answering questions from a different audience each time. The following is an excerpt from the course outline:

> For this group assignment, you will read a newspaper article about a current topic of your choice and create and give a poster presentation with your partners. The purpose of this presentation is two-fold: (1) to explain orally why you chose the topic and what you learned and thought about it and (2) to answer any questions that your audience may have after listening to your talk. You will have three chances to do your presentation, but at each round, you will talk to a new group of three or four classmates. Think of this as an opportunity to improve your oral performance. (Course syllabus)

The first semester presentations took place in July, which meant that the students had known each other for about three months. These tasks were performed over two lesson periods and in several concurrent sessions. Like the other groups in the class, our focus group gave their presentation to a small audience. After 20 minutes, the audience was asked to move to another presentation, and the focus group then welcomed a new audience. In this way, they had the chance to repeat the task three times in a single lesson (see Figure 1). This poster carousel task was intended as a stepping-stone toward the Semester 2 task, which involved a similar presentation but on a one-off basis to a single audience (whole class).

In short, students repeated the same presentations three times in a single lesson in the first semester. There was a five- to six-minute interval between each round of presentations, which was intended mainly for the audience to fill out a feedback form and move to the next poster. In order to keep the presenters focused, during that time, they were simply advised to think about how to improve their presentations but were not given any other specific instruction as to how to use the time.

The project involved two types of repetition, exact task repetition and procedural repetition (a distinction made by Kim, 2013). The former type refers to students repeating the same presentation three times whereas the latter refers to their interacting with a different audience at each round during the five-minute question-and-answer session that took place immediately after their presentations. As such, the discourse of the former phase was mostly planned while the discourse of the latter phase was contingent on the audience's questions and comments. As we will see later, within this lesson design, students were led to work repeatedly on the problem of how to express the same material. In other words, there was a trajectory through the task that involved reiteration and rearticulation.

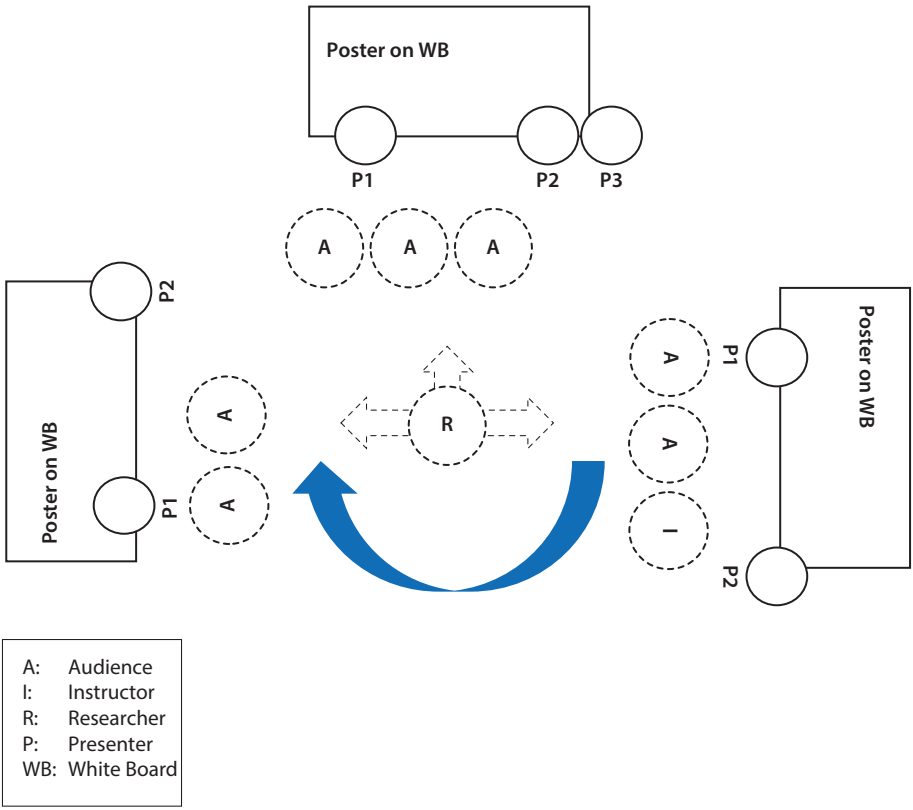

Figure 1.

Table 2. Characteristics of the poster presentation tasks

	Semester 1 Task
Group	2 or 3 people
Content	Any current topic in the news
Length	20 minutes including Q&A
Materials	Posters and any other visual aids
Preparation mostly done outside the classroom	Reading and discussing the chosen article, making a poster, and preparing scripts
Number of performances	3 times in a single lesson
Audience	2–4 people

Data collection

This chapter focuses on the students' task-related classroom discourse[1] including the discourse of the three OPs and question-answer sessions and the presenters' backstage interactions between rounds. These data were audio- and video-recorded. Where useful, students' journals and retrospective interview data as well as the researcher's field notes are also drawn on. The first author was the instructor of the course[2] and participated in the poster presentations as an audience member. This triple role as teacher, researcher, and audience member only allowed her to pay attention to one group and thus prevented her from observing other groups. To compensate for this, the present study involved the second author (henceforth referred to as the researcher) who, by default, quietly watched and listened to participants' interactions during their poster presentations.

Data analysis

In this research, we conducted two types of discourse analysis. First, we used an ethnography of communication (EC) approach to examine communicative patterns in the focal classroom in which students' task performances were situated (Duff, 2002; see Appendix for transcription conventions). According to Saville-Troike (2003), EC seeks to identify what a newcomer to a community must know so as to communicate appropriately in that community. To this end, it provides "an analysis of *speech events*, *activities*, or *tasks* as crucial sites for learning, speaking and performing" (Duff, 2002, p. 292, emphasis original). Thus, EC seems a potentially productive methodology to use given that an important focus of the present study was formed by activities in which L2 students participated as they interacted with others in their classroom.

Second, following Putney et al. (2000), we compared the three performances of each pair/group and traced the students' participation in the poster presentations and associated backstage interactions by moving back and forth in time to trace the interactional trajectories that led to changes in students' performances. In this process, we identified what Newman, Griffin, and Cole (1984) called *tracers* to track students' learning across task-related events. This analytical concept was explained in their research as "a bit of knowledge or some procedure that was taught the children in one of

1. The larger study of which this is a part involved multiple data sources including audio- and video-recorded task-related interactions, the researchers' fieldnotes and journals, audio-recorded interviews with students, and the collection of relevant documents (such as peer feedback forms, students' notes and presentation materials).

2. This might raise questions of an ethical nature about possible coercion to participate in the study (see Duff & Abdi, 2016, for a relevant discussion). To reduce any power differential, we asked the students to submit their consent forms to the administration office one week after the initial information session. We then obtained a final consent from the students after their course grades were submitted.

the settings, which was potentially useful if they recognized that they were being confronted with what they considered that same task in the new setting" (pp. 174–175). So tracers in this study included concepts, words, and phrases that were talked about in the presenters' interactions with their audience and group members during and between their performances.

Teacher expectations and associated practices

Before turning to the individual cases, it is worthwhile examining some of the cultural resources that might have mediated the students' orientation toward and engagement in the presentation task. Given that socialization entails "appropriation of culturally valued goals and intentions" (Ushioda, 2008, p. 25), the teacher's expectations about the presentation task are of particular relevance here. One major feature of a "good" presentation was cooperation among group members. The following excerpt shows how the instructor explained this expectation to her students:

Excerpt 1. The instructor's expectation about the task

Inst: Please keep it in mind (0.8) that this is a – group assignment. Which means (0.6) I expect you to work <u>together</u> (0.7) as a group. (1.0) Mm – like I said in April (0.5) I'd like you to learn together – and learn from each other. (1.2) I've seen some group presentations umm where one member talked about something, (0.8) another member talked about something else, (0.8) a:nd the other – talked about something else. And I couldn't see <u>any</u> strong connections among their parts. To me (0.5) what they did (0.6) was to- to give <u>three</u> short presentations – in a time slot. For this assignment (0.7) I'd like you to cooperate – and uh show me (0.5) and your classmates – that you actually worked together.

(June 16)

As indicated by this excerpt, students were expected to work closely with each other and demonstrate their cooperation and cohesion in their task performance. This is reflected in the student transcripts. For example, Ako[3] presented an outline of her group's speech plan at the beginning of their semester presentation.

Excerpt 2. Presenting an outline

Ako: Good morning everybody. **We**'re going to talk about tobacco. What do you think of tobacco. First **Yui** – will talk about tobacco's bad effects – of human beings. After that – **I**'ll talk about regulations about tobacco – non-smoking areas and packages of

3. All the names are pseudonyms.

> tobacco in the world. Then – **Hina** will talk about Japanese recent tendency of tobacco and advantages of – quitting smoking. And **we'll** try to answer all your questions after the – presentation. Now let's start.
>
> (July 21)

By using several time markers (i.e., "first," "after that," "then," "and"), Ako described her group's planned actions in a sequential order. Her use of these linguistic resources evidences framing which seemed to have contributed to the discoursal sophistication (Samuda & Bygate, 2008) of her group's task performance. However, this is not the only act that Ako is performing here; by using the first person plural pronoun (we) and describing her partners' future actions, she is demonstrating her knowledge about what her partners have to say and the way they are to participate in what follows, thereby suggesting that her group worked together to undertake the OP task. In other words, Ako's language in this relatively short text *indexes* (Ochs, 1996) the cooperative work in which the group engaged behind the scenes. This too reflects the expectations that the teacher had made explicit.

Another important expectation promoted by the teacher concerned interaction between the presenters and audience. This is expressed by the teacher when she says: 'When you're presenting, you should umm listen actively to your audience and answer their questions – and uh when you're an audience member, – you should listen actively to the presenters (0.6) to ask questions' (June 16). Relevant to this is the discursive practice of *revoicing* or "a particular kind of reuttering (oral or written) of a student's contribution – by another participant in the discussion" (O'Connor & Michaels, 1996, p. 71), which also helps create cohesion (Duff, 2000).

There were two major kinds of revoicings in this classroom community. The first is the audience's revoicing of presenter contributions. For example, the instructor as an audience member often revoiced student presenters' utterances to clarify their meaning and to suggest more grammatical and linguistically complex ways of constructing what she understood to be the students' intended meaning (see Mohan & Beckett, 2003, for a relevant discussion). This type of revoicing overlaps with the concept covered by the term recast[4] (see, for example, Lyster, 1998). The other kind of revoicing is the presenter's restating of audience contributions. From her experiences with previous groups of students, the instructor had found that many students seem to focus mainly on the delivery of their prepared talk and put less emphasis on the question-and-answer session. To improve this situation, the instructor would often ask presenters to repeat their audience's questions during the question-answer period with the goal of promoting interactions between the presenters and the audience and

4. In fact, Mohan and Beckett (2003) refer to this as a 'functional recast'.

helping the students learn this discursive practice. In what follows we consider case by case the data for individual students in the group.

Ryo's continued effort to improve his speech

The first case is Ryo, who arguably made the most of his repeated experiences with the OP task. He worked with two female classmates (Moe and Shiho) and decided to discuss online pet shopping, so his data includes turns from his colleagues. As the final speaker of the group, Ryo discussed the merits and demerits of online pet shopping. He was unique in that, unlike his classmates, he did not prepare a script; rather, he spoke from the poster which contained photos of animals and a table (Figure 2). As Ryo read only two sentences from the poster, his speech was relatively spontaneous.

A comparison of his three performances (see Table 3) reveals that Ryo made an important change to the final section of his second and third performances. Table 3

Online Pet Shopping Rules	
Merits	**Demerits**
You can buy pets without going to the pets shop. You can find many kinds of animal from many place. You can buy animals much cheaper than in the pet shops.	You can't really sure that it is safe to buy. **It might be lack of information for buyer**.

Figure 2. Extract from Ryo's group's poster

Table 3. Ryo's three performances

Round 1	A:nd (0.6) some of people think it is not safe (0.7) enough to buy (0.6) lots of animal. (1.1) A:nd at last (0.8) it might be lack of informations (0.8) for buyers. (2.1) So there's lots of – merits, a:nd demerits parts. So: please check this (0.4) out.
Round 2	A:nd (0.6) we don't know (0.5) it is really safe to buy (0.5) pets. (1.1) So: (0.6) some of the homepages is not really safe. (0.8) A:nd it might be uh lack of informations for buyers. (1.1) **A:nd I think (0.7) some of dogs (1.3) on the internet (1.7) maybe sick dog (1.2) uh: (1.1) they have something (2.3) weird, (0.5) or (0.5) maybe yeah very weird, (0.7) and they may die soon**. So we don't know lots informations. (0.6) Uh we don't know (1.1) it is yeah (0.6) it's (0.6) safe to buy animals. (1.5) So: (0.5) I think (1.6) we should check it more carefully.
Round 3	A:nd (0.5) we don't know – it is really safe to buy – pets. (0.6) So: they are not really safe. (0.5) A:nd it might be a lack of informations for buyers. (0.5) **A:nd I think – some of dogs on the internet (0.6) may be sick. I mean (0.5) they may have (0.7) disease or uh: (1.1) mm (0.6) parasite. (1.0) And they may die soon**. (0.5) We don't know lots of information. (0.6) So we don't know (0.6) it's safe to buy animals. (0.7) So: I think (0.7) we should check it more carefully.

juxtaposes the excerpts from his three performances. Bold type is used to signify new information that Ryo added to his speech. To learn how this change took place, we examined not only what Ryo said in each round, but also what transpired between the three task performances.

Excerpt 3 is the last part of Ryo's first performance where he discussed the problems of online pet shopping.

Excerpt 3. Round 1

```
Ryo:  A:nd (0.6) some of people think it is not safe (0.7) enough to buy
      (0.6) lots of animal. (1.1) A:nd at last (0.8) it might be lack of
      informations (0.8) for buyers. (2.1)
      So there's lots of – merits, a:nd demerits parts. So: please check
      this (0.4) out.
```

(July 14)

The message can be summarized as follows:

> Some people think that it is not safe to buy many types of animals online. And there might be a lack of information for buyers. There are many merits and demerits. So we should consider the merits and demerits of online pet shopping.

This is followed a few minutes later in the question-answer period, where one of Ryo's partners, Moe, invited the researcher, who happened to be observing her group's presentation, to ask questions. The researcher later reflected on this event in his field notes as follows:

> After the lesson, I told Moe that I was a bit surprised that she invited me to ask a question. She said with a big smile on her face that since none of the audience members seemed ready to ask a question, she wanted to give them more time to think.
>
> (July 14, 2005)

Thus, Moe drew upon the researcher's ability in order to address what she perceived to be her audience's unreadiness to ask questions.

Moe's nomination of the researcher as the first questioner (see lines 1–3 in Excerpt 4 below) led to the following exchange: in line 4, the researcher asks the presenters if a buyer is guaranteed a complete refund in case his or her ordered pet dies during transportation or shortly after its arrival. Ryo and Moe answer that they do not think that the buyer will get a complete refund, in lines 5 and 6, respectively. In line 7, the researcher provides the gist of the discussion so far.

Excerpt 4. Round 1 Question Time

```
1  Moe: ((to the researcher)) Do you have any question? (1.4) ((looks at
        the researcher))
2  Res: Oh – me?
```

3 Moe: (0.6) Yes. ((laughs))
4 Res: Sure. ((laughing)) (0.8) Okay. Umm (0.5) **when you (0.5) buy an animal from an online sho- uh: pet shop (0.5) they send you the pet that you ordered, right? And if (0.6) he or she dies – on his or -her way (0.6) or right after his or her arrival, do they guarantee a complete payback? (0.9) I mean – will you get all of your money back, if the animal dies**?
5 Ryo: I don't think so. Mmm.
6 Moe: I don't think so. =
7 Res: = So it's not guaranteed.
8 Ryo: So- some of (1.1) uh: (0.7) online – pet shop – are (0.6) guarantee but – most of (1.2) online pet shop doesn't have.
9 Res: Okay. So maybe that's another reason – why you may not want to do online pet shopping?
10 Moe: Mm. Yeah.
11 Res: **It's another demerit, then.**
12 Ryo: Mm. ((nodding))
13 Moe: **Demerits. I- we can't- we:: can't (0.9) know the health, health of- of a pet.**
14 Res: I see.

(July 14)

In line 8, Ryo elaborates on the point by saying that only a few shops guarantee a complete refund. The researcher then again summarizes the discussion in line 9, to which Moe indicates her agreement in line 10. In line 11, the researcher further states that it is another demerit that most online shops do not guarantee a complete refund. Hearing this statement, Ryo indicates his agreement by nodding in line 12. In line 13, Moe further elaborates on the point by saying that buyers cannot know the animals' health. Importantly, the researcher played an active role in transforming the event.

After the first presentation round, Ryo and his partners talked in Japanese about their own performance. Excerpt 5 is taken from this backstage talk (Italics are used to signify L1 utterances). As mentioned earlier, the major medium of instruction was English. Thus students' use of L1 was an unofficial part of the lesson.

Excerpt 5. Backstage talk between Rounds 1 and 2

1 Ryo: *And should I talk about that? I mean that question about the possibility of the animals' dying during transportation.*
2 Moe: *That's a* demerit, *right*? (0.6) [*Yes. Sounds good.*
3 Shiho: [*Yes. I think that's a good idea.*
4 Ryo: OK.

(July 14)

In line 1, Ryo asks Moe and Shiho whether it would be better if, in his second performance, he addressed the point raised by the researcher (i.e., the possibility of ordered pets dying during transportation) earlier. In their subsequent contributions, both Moe and Shiho indicate an agreement to this suggestion. Note that the point raised in the discussion was initially about whether or not a complete refund was guaranteed in case of the pet's death. But the students apparently shifted their focus to the possibility of the pet's death itself.

In his second performance, as indicated by bold type, Ryo added new information (i.e., the possibility of the pet's death). However, he seems to be having difficulty expressing it, which is indicated by the relatively long pauses (e.g., 1.1 seconds, 2.3 seconds) and the vague expression (i.e., *something weird*).

Excerpt 6. Round 2

Ryo: A:nd (0.6) we don't know (0.5) it is really safe to buy (0.5) pets. (1.1) So: (0.6) some of the homepages is not really safe. (0.8) A:nd it might be uh lack of informations for buyers. (1.1) **A:nd I think (0.7) some of dogs (1.3) on the internet (1.7) maybe sick dog (1.2) uh: (1.1) they have something (2.3) weird, (0.5) or (0.5) maybe yeah very weird, (0.7) and they may die soon.** So we don't know lots informations. (0.6) Uh we don't know (1.1) it is yeah (0.6) it's (0.6) safe to buy animals. (1.5) So: (0.5) I think (1.6) we should check it more carefully.

(July 14)

Despite the hesitations and pauses, Ryo has significantly elaborated his contribution to the presentation, drawing on both the previous question-answer phase and the group's backstage talk. After their performance, Ryo and his partners had another talk in Japanese (Excerpt 7 below). In line 1, Ryo asks his partners how to say "*byooki*" (sickness) in English. Shiho then suggests two word choices (*sick* and *ill*) in line 2. Ryo implicitly rejects both of them by asking for other choices (line 3). In line 4, Moe suggests a new word (*disease*), which is accepted by Ryo in the following line. In the same line, Ryo asks another question: "*how about kiseechuu* (parasite)?" Unfortunately, neither Moe nor Shiho knows how to say this in English. In line 11, Ryo consults his dictionary and finds the English word.

Excerpt 7. Backstage talk between Rounds 2 and 3

1 Ryo: *You know what – how do you say sickness in English*?
2 Shiho: Sick *or* ill?
3 Ryo: *Mm – any other words*? (3.2)
4 Moe *How about* disease?
5 Ryo: *Ah: right,* disease. *And what was it*? *Then* (0.5) *how about* ***parasite***?

6 Moe: *I don't know*. (0.7) *How do you say that*?
7 Shiho: *I don't know*.
8 Ryo: *Last time I didn't know how to say that in English, so I said* **"something weird."** ((laughs))
9 Shiho: ((laughing)) *(Disease is) like* ***something weird***?
10 Moe: *Right – it sure is* ***something weird***. ((laughs))
11 Ryo: ((consults his dictionary.)) (16.3) *Found it. It's* **parasite**!
12 Moe: AH: **parasite**!
13 Ryo: *Yeah*.
14 Shiho: AH: – **parasite** – *right*.

(July 14)

In short, in this second phase of backstage talk, initiated by Ryo's questions, the group worked together to come up with appropriate words to express Ryo's intended meaning. It is important to note here that Ryo's utterance in line 8 suggests that as a result of his first task performance, he noticed what he could not say exactly in English (Swain, 1995). The next extract, Excerpt 8 below, shows how his talk evolved in the ensuing final presentation.

Excerpt 8. Ryo's Round 3

Ryo: A:nd (0.5) we don't know – it is really safe to buy – pets. (0.6) So: they are not really safe. (0.5) A:nd it might be a lack of informations for buyers. (0.5) **A:nd I think – some of dogs on the internet (0.6) may be sick. I mean (0.5) they may have (0.7) disease or uh: (1.1) mm (0.8) parasite. (1.0) And they may die soon**. (0.5) We don't know lots of information. (0.6) So we don't know (0.6) it's safe to buy animals. (0.7) So: I think (0.7) we should check it more carefully.

(July 14)

Although he produced relatively long pauses (1.1 seconds and 0.8 seconds) and hesitation markers (*uh*:, and *mm*), Ryo successfully incorporated into this third performance the two words (*disease* and *parasite*) that the group had talked about earlier. This, then, was a case where students acted in a sustained way upon the researcher's feedback to improve their subsequent task performance.

It is worth repeating that Ryo was the only student in the class who did not prepare a script for the presentation. Having observed his performance, many of his classmates gave positive feedback especially on his oral fluency and use of non-verbal cues such as eye contact. Interestingly, his partner Moe wrote in her written reflection submitted a week later as follows:

> The most shocking thing was Ryo presented with no paper. His tone was nice. But on the other hand, I felt bitter about that. At that time, I found my goal. My goal is read with no paper and talk nice. He serve as a stimulus to me.

By working side by side with Ryo, Moe realized the importance of not depending too much on her script and, as a result, came to see him as a peer role model in that respect. Importantly from that point, Moe used cards containing main points and key phrases instead of writing out her speech. Thus, she benefitted from observing the repeated performances of her partner.

Nagi's unpacking of a complex sentence

Table 4 shows that Nagi's speech remained virtually the same across the first two performances, but it changed remarkably in the third round.

Table 4. Nagi's three performances

Round 1	For example, last year, Tokyo's temperature -didn't go down so much. This is called umm heat island phenomenon. Heat island is- (0.6) umm heat island phenomenon is a – phenomenon, in which urban cities are hotter than – rural areas, because – of the "concentrated development and massive energy consumption."
Round 2	For example, last year, Tokyo's temperature didn't go down so much. This is called heat island phenomenon. Heat island phenomenon is a phenomenon, in which urban cities are hotter than rural areas because of the "concentrated development and massive energy consumption."
Round 3	For example, last year, Tokyo's temperature didn't go down so much. This is called heat island phenomenon. **Because of heat island phenomenon, urban cities are hotter than rural areas, and Tokyo is (0.6) hotter, (0.7) because it is crowded with – people a:nd buildings, and – many business, and they need uh: a lot of energy**.

Excerpt 9 illustrates how the instructor helped Nagi and her partner Jun to unpack nominalized phrases and understand what they meant, and demonstrated how to express the same content in a more audience-friendly manner. This interaction occurred right after the students finished their second round performance. The instructor draws the presenters' attention to their explanation of the heat island phenomenon (line 1) and asks them to explain it again (line 3). Nagi then quotes their article verbatim (line 4). After acknowledging this response, the instructor asks a more focused question about two nominalized phrases (line 5). In lines 6–10, the presenters "unpack" the complex phrases with the instructor (see Mohan, Slater, Beckett, & Tong, 2015, for a relevant discussion), who then summarizes their discussion (line 11).

Excerpt 9. Round 2 Question-and-Answer Time

1 Inst: Thank you for your informative presentation about the Cool Biz Campaign. You did a good job explaining what the campaign is about and how heat waves affect our life. Umm could we go back to: – what you're saying about **the (0.5) heat island phenomenon**? (0.6)
2 Nagi: [Yes.
3 Inst: [What did you say?
4 Nagi: Umm (0.6) well – **heat island umm phenomenon, (4.3) is a "phenomenon is which urban areas are hotter than rural areas due to concentrated development and massive energy consumption."** ((reading her notes))
5 Inst: Oh ok. I see. (0.5) What did you mean by (0.4) **concentrated development and – what was it, uh – massive energy consumption**?
6 Nagi: umm – many people and building,
7 Inst: Mmm-hmm, (0.8)
8 Jun: Tokyo is crowded, so: (0.9)
9 Inst: I see. It's crowded with people and buildings.
10 Nagi: so umm need a lot of – energy.
11 Inst: **So if I understood you correctly, (0.6) because there're – so many people and buildings, and also because umm – businesses and industries and schools, so many things have developed in urban cities – like Tokyo, a lot of energy is being used.** – Is that right?
12 Nagi: [Yes yes.
13 Jun: [Yes
14 Inst: I see.

(July 14)

In her teaching journal, the instructor gave the following reasons for pushing Nagi to paraphrase the original sentence: "I was not sure if the audience understood the complex expressions that she appeared to have borrowed from the article. I also wanted to see if they (Nagi and Jun) knew what they were talking about" (July 14). Thus, this was a purposive act of the instructor to socialize the presenter into the discursive practice of paraphrasing.

Excerpt 10 shows how Nagi and Jun engaged backstage in languaging in order to make sense of some phrases and change them into more audience-friendly utterances on their own. In line 1, Nagi initiates their discussion about the heat island phenomenon. As her utterance in line 3 indicates, her focus is on the two nominalized phrases that they unpacked in the question-and-answer session. In line 6, Jun tells Nagi in Japanese what each phrase means, which is followed by Jun's summary of the original

text. Having achieved a shared understanding with Jun, Nagi starts to express the content in English (line 9). The last part of this utterance is recast by Jun (line 10), which results in Nagi's incorporation of the *because of* connector in line 11. In the same turn, Nagi goes on to ask Jun to help her express her intended meaning in English. In response to this call, Jun consults her dictionary and comes up with an answer, informing Nagi that the adjective *crowded* should be followed by the preposition *with*. In line 13, Nagi verbalizes a full sentence, which is not error-free but nonetheless expresses the original meaning of their article without using nominalizations.

Excerpt 10. Backstage Talk between Rounds 2 and 3

1 Nagi: *As for the* heat island *(phenomenon), what should I say*?
2 Jun: *What*?
3 Nagi: **"Con- Concentrated development and – massive energy consumption."**
4 Jun: *Oh I see. That*.
5 Nagi: *Yeah*.
6 Jun: *Umm* – **concentrated development *means crowded, and* massive energy consumption – *means a great amount of energy is used*.**
7 Nagi: *So (0.8) we can say something like – Tokyo was hotter than other cities due to the heat island effect because – it's densely populated and filled with buildings and* (0.9) *so a great deal of energy is used especially in the summer*.
8 Jun: *GRE:AT!. Sounds good!*
9 Nagi: *Then* (1.6) **Tokyo was (0.6) more hot- *oh no* hot- hotter (0.4) because (0.6) the heat island phenomenon,**
10 Jun: **because of, – heat island phenomenon**?
11 Nagi: *Right*. (0.5) **Because o:f heat island phenomenon,** (1.4) **Tokyo is crowded,** *how do you say crowded with people and buildings*.
12 Jun: *Just a moment*. ((consults her dictionary)) (20.9) *Oh:: you need* with, *it seems*. (0.6) **Crowded <u>with</u> people.**
13 Nagi: *Ok*. **Tokyo i:s (0.5) crowded with people and (0.6) buildings, and need a lot of energy (0.9) in the summer.**
14 Jun: *Sounds good!*

(July 14)

In short, Nagi's case is an example of not only language learning but also learning the content (concepts of heat island phenomenon/energy consumption) better.

Ako's appropriation of audience questions through procedural repetition

Ako's case illustrates how procedural repetition provided her with a linguistic affordance to improve her subsequent task performance. She worked with two female classmates (Yui and Hina) to make a presentation on the effects of smoking. Like many other classmates, she used her script in her presentation. Because of this, her speech remained virtually the same over the three performances. However, as the following excerpts will show, the question-answer period after her second performance seemed to have helped Ako to better prepare for her performance in the third question-answer period.

A few minutes into the question-answer period, an audience member (Miki) asked the presenters what they would do if someone were smoking in the street (lines 1 and 3). Ako's answer to this question was that she would pretend to cough and walk quickly so as to make the smoker realize that it is bad manners to smoke in the street (line 4). Then, Miki summarizes Ako's previous utterances by saying "You're saying something. You're sending signal" (line 5), to which Ako answers with a simple yes (line 6). As will become evident in the next excerpt, Ako appropriated the word 'signal' in her subsequent performance.

Excerpt 11. Round 2: Answering an audience question

1 Miki: **(If you) don't like smoking, (0.6) and (0.8) there's (one) people who (1.1) is smoking uh in the sidewalk, (1.2) and- a:nd (2.1) you don't want to:: (0.5) smell the smoke, (1.1) and (3.1) you want to say something to (0.9) he or she. (1.2) What would you say to them?**
2 Ako: Uh: (1.5)
3 Miki: In store you can say that (1.7) there're- customers. (1.1) Because sidewalk is very dangerous. (2.0) Some of people are (xx[x)
4 Ako: [**Uh: I often (1.5) pretend to cough, (1.0) and (2.8) walk through (2.5) quickly. (2.1) So uh: (0.9) I try to realize (1.2) to be umm, (0.7) I make (0.7) him or her to realize (0.5) that is bad manner.**
5 Miki: **So you're saying – something- you're sending (1.5) signal.**
6 Ako: Yes. ((smiling))

(July 21)

Excerpt 12 comes from the question-answer session of the group's third poster presentation. Having finished her group presentation, Ako first invites the audience to ask questions. However, no one volunteers. Ako then announces that she will ask a question and asks the audience what they would do if someone were smoking in the street. Although using different wording, Ako is using the very question that she received from her previous audience. She then invites one student (Sae) to answer the

question in the same line. In line 5, another student (Haru) self-nominates and gives her an answer. Finally, Ako answers her own question, appropriating the word "signal" from her previous audience (line 6).

Excerpt 12. Round 3: Appropriating an audience question

1 Ako: Are there any other questions? (12) ((looks around the classroom)) Ok, (0.6) I will ask a question. (2.2) **If you don't smoke, (0.5) and you don't like the (0.4) smell, what will you do (0.6) if someone is smoking on the- (0.8) in the street. (0.8) How about you**?
2 Sae: (0.8) Mm – I ignore.
3 Ako: You will say nothing?
4 Mai: Mm (0.6) nothing.
5 Haru: *Here*. ((raising her right hand)) I will tell the person (0.7) this is non-smoking. (0.6) So please – stop smoking. ((students laugh))
6 Ako: *OKay*. (1.2) **Uh: I often pretend to cough – and walk quickly. It's – um my – signal**. ((students laugh)) **I want to make him to realize, (0.7) make him to realize that is a bad manner**.

(July 21)

What is interesting about this example is that Ako appropriated a question from her previous audience and used it to involve her new audience in the subsequent performance. Her utterances produced in this performance contain as many grammatical errors as those produced in the second performance. However, she made fewer pauses, and fewer hesitations, reformulations and repetitions, thus sounding more fluent.

Rena's non-appropriation of audience feedback

For the poster presentation, Rena worked with another female student Mao and talked about the Japanese government's "cool biz campaign." Like Ako and Nagi and unlike Ryo, Rena used her script in her presentations. In fact, her goal was to "read her script well" (written reflection). As such, her three performances were virtually identical except where the instructor asked questions. Excerpt 13 comes from Rena's group's second performance.

Excerpt 13. Round 2: Instructor's questions during the presentation

1 Rena: …as all of you know – the environmental problem – is getting more – and more serious today. One of the biggest problem is – greenhouse effect. It is caused by the increase of greenhouse gases. (0.7) Such as car- carbon dioxide, methane and Freon. If the greenhouse effect continues in future, – they predict – that

surface of the sea water level will rise about one centimeters by (now) – twenty first century…When you watch it on global scale – many islands will set in a sea by surface of the sea rises. (0. 6) In addition – areas in (xxx). Shortage of water and the drought – will come out.

2 Inst: **So the greenhouse effect is said to cause global warming – right**?
3 Rena: (0.8) Global warming? (0.7)
4 Inst: Yes. Global warming.
5 Rena: Ah: ((nodding)) yes yes.
6 Inst: **And global warming can cause other umm – serious problems – including a rising sea level and umm abnormal weather events like droughts and floods – right**?
7 Rena: (0.5) Yes.

(July 21)

In line 2, the instructor summarizes Rena's previous utterances and invites her confirmation by using a question tag ("right?"). Importantly, the instructor is introducing the term global warming. In line 3, Rena simply repeats part of the instructor's previous utterance. The instructor then indicates that Rena has heard it right (line 4). In line 5, Rena shows her understanding of the term. In line 6, the instructor further recasts Rena's speech by suggesting a more compact statement (see Mohan & Beckett, 2003, for a relevant discussion). During the question-and-answer session, the instructor asked two more questions, which Rena and Mao could not answer. The examination of the transcripts showed that neither of them changed their speech, sticking to their original plan. Rena later said in her interview:

> I didn't know the answers to the teacher's questions about global warming… Even if I had known the answers, I would probably have not changed my speech because I was afraid of making grammatical errors and failing to perform smoothly. I need some time to write English sentences. (Japanese in original) (interview, July 26)

Similarly, when asked at the interview how they spent the time between the rounds, Mao answered, "*I practiced my part by reading the script and Rena did the same*." (July 27). This was evidenced by the researcher's observation of their backstage behaviors, which included individually reading their scripts in a soft, almost inaudible voice, and consulting their dictionaries to check pronunciation of some words.

In sum, Rena did not change her second or third performance for two reasons. First, she and Mao simply did not have enough knowledge to answer the last two questions. Second, their major concern was to avoid risk taking and to perform accurately and fluently. In other words, they were more *performance-oriented* than *learning-oriented* (Dweck, 1989). However, Rena's comment suggests that she might have acted upon her audience feedback had she had more time between performances.

Discussion

This study has focused on three iterations ('rounds') of a classroom task, each of which was made up three phases: the oral presentation, an ensuing question-and-answer session, and private 'backstage' talk among the group members. The data show how the participants converted the task (or 'task-as-workplan') into distinct interpersonal activities ('task-as-process'). In particular, the analysis has tracked how through the iterations, appropriation has taken place by processes of noticing associated with languaging and revoicing. From this perspective, the iterations ('rounds') of the task and its different phases provide a context for learners to generate their own appropriations through the affordances that they perceive. More specifically, the analysis showed that task repetition provided the presenters with important opportunities for noticing and languaging. Between their performances, the focal students and their partners talked briefly about the language and content of their presentations in Japanese. After their first performance, Ako, Nagi and Ryo discussed with their partners some of the points raised by their previous audience and decided to address them in their subsequent presentations. For instance, in his second performance, Ryo had difficulty expressing the new content in English and ended up using vague expressions. However, this experience allowed him to "notice a hole" (Doughty & Williams, 1998) in his L2 productive repertoire. Dissatisfied with his own performance, Ryo, before his third performance, talked with his partners about the two words he had difficulty expressing in English (i.e., *disease* and *parasite*), again using Japanese. As a result of this L1 backstage talk, Ryo was able to express his intended meaning more precisely in English on the third round. Likewise, Nagi and her partner engaged in L1 backstage talk to unravel the meaning of the linguistically dense phrase from their chosen article, which made her explanation of the heat island phenomenon more audience-friendly.

These examples suggest that task repetition provided the presenters with multiple opportunities for languaging in the form of collaborative L1 dialogue that helped them (1) familiarize themselves with their intended meaning, which in turn might have allowed them to attend to language without focusing too much on content, (2) unpack linguistically complex sentences and communicate their meaning in more audience-friendly language, and (3) make a decision about the language and content of their speech in a relatively short period of time (see Kobayashi, 2004, for similar findings). Thus, the analysis of the presenters' backstage interactions coupled with the analysis of their public performances has provided an important window into how their primary focus shifted from meaning to form through task repetition. This finding not only corroborates the findings of experimental studies (e.g., Bygate, 2001, Fukuta, 2016), but also points to the need to examine how students' backstage engagement with the same material may shape their subsequent task performance. Additionally,

task repetition provided Moe with multiple opportunities to work side-by-side on the same task with Ryo, from whom she learned how to do an OP.

The analysis also demonstrated that although students repeated the same task three times, each of these performances generated a different activity (Coughlan & Duff, 1994), indicating the co-constructed nature of presenters' L2 learning through task repetition. In the first case presented in this chapter, for example, it was Moe's nomination of the researcher as the first questioner in the first round that led to Ryo's subsequent noticing and languaging. Also, student agency played an important role. Although they were given no specific instruction, most students in this study used the break between the three rounds to reflect on and refine their performances, which provided different affordances to different students. Ryo and Nagi perceived their own problems through their interactions with the audience (i.e. the instructor and researcher) and jointly created a context of learning with their partners by languaging about the language of their speech. Likewise, to invite her audience's contributions, Ako used one of the questions that she had received from her previous audience.

Relevant to this point is Bakhtin's (1986) notion of *addressivity*. According to Lantolf and Thorne (2006),

> When an individual produces an utterance, this instance of a communicative gesture consists of drawing from prior voices and articulations emerging from that community, tailoring the utterance so that it may achieve the speaker's goals, and anticipating the potential responses from the interlocutor or community to the utterance in question. (p. 239)

Similarly, Ako's production of the question entailed drawing from prior voices and articulations in her classroom, tailoring the utterances so as to achieve her goal, and anticipating the responses from her audience. This was an agentive act that Ako integrated in her task performance. In short, the data suggests that as "active agents of information processing and action" (Hatano, 1993, p. 156), the focal students *acted beyond the structure of the task* (Heath, 2000) by creating contexts for backstage learning to prepare for their upcoming public performance. Their sustained purposeful engagement across events seems to have been enabled by the combination of student agency and task repetition that were situated in a particular classroom context.

These findings suggest that audience feedback, especially, teacher feedback, might be an important part of learning through repeated engagement in the OP task. However, students in the present study did not have equal access to teacher feedback as the instructor could attend only one presentation at a time. While the pair/group that received teacher feedback on their first performance had two chances to make changes or incorporate new elements into their speech, the pair/group that received teacher feedback on their third performance did not have such a chance. It would be necessary for teachers to consider how to provide more equal feedback to their students.

Also, several students mentioned that they did not change their second or third task performances by using their audience's questions or comments. They said that their primary concern was to deliver their speech in grammatically accurate language and in a timely fashion, so they opted to stick to their original speech plan to avoid making errors. In fact, many students had memorized prepared scripts. For instance, Rena mentioned that she did not know the answers to their audience's questions and thus needed to find the relevant information and plan how to express it. To encourage students to act on their audience's comments and questions so as to improve their task performance, teachers could give them short planning/reflection time between performances together with explicit instruction as to how to use this time. In other words, students in their groups could be guided to reflect jointly on their task performances and plan how to express their meanings in the target language. Such guidance might promote group interaction which is considered to be instrumental in ensuring access to and active engagement with linguistic affordances (van Lier, 2004).

Furthermore, to allow time to search for information and make substantial revisions, teachers could give a longer period of time (e.g., one week) before the final performance. In other words, the delayed repetition of an OP task could be used to encourage students to make better use of audience comments. In addition, like Kobayashi (2016) suggests, students could be encouraged to take notes of points raised by their audience and address them in their written reports. We consider such an approach as important in terms of academic discourse socialization because it could afford students opportunities to engage with the same content in a written mode by repacking the nominalizations that they once unpacked through task-related interactions with their co-participants (see Mohan & Beckett, 2003; Mohan et al., 2015, for a relevant discussion).

Conclusion

The tasks repeated in this chapter involved both planning and delivery of presentations, as well as improvisation in the following question-answer sessions. Some pairs/groups went beyond this task requirement to use the short break to address linguistic problems that had become evident in their earlier performance. Such responsiveness is deemed essential for L2 students to learn through their interactions and collaboration afforded by task repetition. According to van Lier (1996), learning occurs "when the new is embedded in the familiar, so that risks and security are in balance" (p. 171). Task repetition reported here seems to have served as a type of pedagogical scaffolding (van Lier, 2004) because it provided the security through the repeated format, and at the same time encouraged the students to move toward participation in unscripted, spontaneous forms of L2 talk. However, as van Lier (2008) puts it, learners "must be

assisted and stimulated (in a dialogical process…) when new responses, behaviors or ventures occur that depart from the script in developmentally promising ways" (p. 62). The data have shown how backstage talk, revoicing, and languaging can be valuable vehicles for this. However, future research could also benefit from longitudinal designs that examine how performance-oriented students like Rena and Mao learn to use backstage "interactions as a tool for mediating and assisting learning" (Walsh, 2011) across task-related events.

One clear limitation of the study is that on its own the research methodology used cannot reveal the status and evolution of learners' ZPDs. Also, the present study focused on the repeated performances of the presenters and did not examine whether/how this contributed to students' participation as audience members. Future research could examine the possible impact of procedural repetition on audience members' participation in OPs. Moreover, future work could examine how students decide the focus of their backstage talk when they have limited time but multiple items to address. Which items do they choose? Whose concerns get addressed? Whose suggestions get adopted when there is more than one competing idea to solve the linguistic problem under discussion? Answers to these questions would enrich our understanding of interpersonal processes involved in L2 learning through repeated engagement in OP tasks.

Acknowledgments

We are very grateful to Lori Ann Desrosiers, as well as to the anonymous reviewers and the editors for their constructive comments, which greatly helped us to improve the manuscript. The first author is particularly indebted to Alan Jones and Stephen Moore for their invaluable guidance and feedback on the project from which this chapter originates. Last but not least, we would like to thank our participants for their cooperation.

References

Bakhtin, M. M. (1981). *The dialogic imagination* (M. Holquist (Ed.) & C. Emerson & M. Holquist (Trans.)). Austin, TX: University of Texas Press. (Original work published 1975).

Bakhtin, M. W. (1986). *Speech genres and other later essays* (C. Emerson & M. Holquist (Eds.) & V. W. McGee (Trans.)). Austin, TX: University of Texas Press. (Original work published 1979).

Batstone, R., & Philp, J. (2013). Classroom interaction and learning opportunities across time and space. In K. McDonough & A. Mackey (Eds.), *Second language interaction in diverse educational context* (pp. 109–125). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.34.09ch6

Breen, M. (1987). Learner contributions to task design. In C. N. Candlin & D. Murphy (Eds.), *Language learning tasks: Lancaster practical papers in English language education* (Vol. 7, pp. 23–46). Englewood Cliffs, NJ: Prentice-Hall.

Bygate, M. (2001). The effects of task repetition on the structure and control of language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 23–48). Harlow: Longman.

Bygate, M., & Samuda, V. (2005). Integrative planning through the use of task-repetition. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 37–74). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.11.05byg

Bygate, M., & Samuda, V. (2009). Creating pressure in task pedagogy: The joint roles of field, purpose, and engagement within the interactional approach. In A. Mackey & C. Polio (Eds.), *Multiple perspectives on interaction* (pp. 90–116). New York, NY: Routledge.

Coughlan, P., & Duff, P. (1994). Same task, different activities: Analysis of an SLA task from an activity theory perspective. In J. P. Lantolf & G. Appel (Eds.), *Vygotskian approaches to second language research* (pp. 173–193). Norwood, NJ: Ablex.

Doughty, C. & Williams, J. (1998). Pedagogical choices in focus on form. In C. Doughty & J. Williams, (Eds.), *Focus on form in classroom second language acquisition* (pp. 197–261). Cambridge: Cambridge University Press.

Duff, P. A. (2000). Repetition in foreign language classroom interaction. In J. K Hall & L. S. Verplaetse (Eds.), *Second and foreign language learning through classroom interaction* (pp. 109–138). Mahwah, NJ: Lawrence Erlbaum Associates.

Duff, P. A. (2002). The discursive co-construction of knowledge, identity, and difference: An ethnography of communication in the high school mainstream. *Applied Linguistics*, 23, 289–322. https://doi.org/10.1093/applin/23.3.289

Duff, P. A. (2007a). Beyond generalizability: Contextualization, complexity, and credibility in applied linguistics research. In M. Chalhoub-Deville, C. Chapelle, & P. A. Duff (2007), *Inference and generalizability in applied linguistics: Multiple perspectives* (pp. 65–95). Amsterdam: John Benjamins.

Duff, P. A. (2007b). Second language socialization as sociocultural theory: Insights and issues. *Language Teaching*, 40, 309–319. https://doi.org/10.1017/S0261444807004508

Duff, P. A. (2008). *Case study research in applied linguistics*. New York, NY: Lawrence Erlbaum Associates.

Duff, P. A., & Abdi, K. (2016). Negotiating ethical research engagements in multilingual ethnographic studies in education. In P. I De Costa (Ed.), *Ethics in applied linguistics research: Language researcher narratives* (pp. 121–141). New York, NY: Routledge.

Duff, P. A., & Kobayashi, M. (2010). The intersection of social, cognitive, and cultural processes in language learning: A second language socialization approach. In R. Batstone (Ed.), *Sociocognitive perspectives on language use and language learning* (pp. 75–93). Oxford: Oxford University Press.

Dweck, C. (1989). Motivation. In A. M. Lesgold & R. Glaser (Eds.), *Foundations for a psychology of education*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Fukuta, J. (2016). Effects of task repetition on learners' attention orientation in L2 oral production. *Language Teaching Research*, 20, 321–340.

Gass, S., Mackey, A., Alvarez-Torrez, M. J., & Fernández-García, M. (1999). The effects of task repetition on linguistic output. *Language learning*, 49(4), 549–581. https://doi.org/10.1111/0023-8333.00102

Gee, J. P. (1996). *Social linguistics and literacies: Ideology in Discourses* (2nd ed.). New York, NY: Routledge.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

Goffman, E. (1959). *The presentation of self in everyday life*. New York, NY: Anchor Books.

Hatano, G. (1993). Time to merge Vygotskian and constructivist conceptions of knowledge acquisition. In E. A. Forman, N. Minick, & C. A. Stone (Eds.), *Context for learning: Sociocultural dynamics in children's development* (pp. 153–166). Oxford: Oxford University Press.

Hawkes, M. L. (2012). Using task repetition to direct learner attention and focus on form. *ELT Journal*, 66, 327–336. https://doi.org/10.1093/elt/ccr059

Heath, S. B. (2000). Making learning work. *Afterschool Matters*, 1, 33–45.

Kim, Y. (2013). Promoting attention to form through task repetition in a Korean EFL context. In K. McDonough & A. Mackey (Eds.), *Second language interaction in diverse contexts* (pp. 5–24). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.34.04ch1

Kim, Y., & Tracy-Ventura, N. (2013). The role of task repetition in L2 performance development: What needs to be repeated during task-based interaction? *System*, 41, 829–840. https://doi.org/10.1016/j.system.2013.08.005

Kobayashi, E. (2007). *Task repetition in an EFL classroom: A language socialization perspective* (Unpublished postgraduate certificate dissertation). Macquarie University, Sydney, Australia.

Kobayashi, M. (2004). *A sociocultural study of second language tasks: Activity, agency, and language socialization* (PhD dissertation). Retrieved from <https://open.library.ubc.ca/cIRcle/collections/ubctheses/831/items/1.0078243>

Kobayashi, M. (2016). L2 academic discourse socialization through oral presentation: An undergraduate students' learning trajectory in study abroad. *Canadian Modern Language Review*, 71, 95–121.

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford: Oxford University Press.

Lynch, T. (2018). Perform, reflect, recycle: Enhancing task repetition in second language speaking classes. In M. Bygate (Ed.), *Language learning through task repetition* (pp. 201–231). Amsterdam: John Benjamins.

Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, 4, 221–250.

Lynch, T., & Maclean, J. (2001). 'A case of exercising': Effects of immediate task repetition on learners' performance. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogical tasks: Second language learning, teaching and testing* (pp. 141–162). Harlow: Pearson Education.

Lyster, R. (1998) Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning*, 48(2), 183–218. https://doi.org/10.1111/1467-9922.00039

Mercer, N. (1992). Culture, context, and the construction of knowledge in the classroom. In P. Light & G. Butterworth (Eds.), *Context and cognition: Ways of learning and knowing* (pp. 28–46). Hillsdale, NJ: Lawrence Erlbaum Associates.

Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach.* Abington: Routledge.

Mohan, B., & Beckett, G. H. (2003). A functional approach to research on content-based language learning: Recasts in causal explanations. *The Modern Language Journal*, 87, 421–432. https://doi.org/10.1111/1540-4781.00199

Mohan, B. A., & Marshall Smith, S. M. (1992). Context and cooperation in academic tasks. In D. Nunan (Ed.), *Collaborative language learning and teaching* (pp. 81–99). Cambridge: Cambridge University Press.

Mohan, B. A., Slater, T., Beckett, G. H., & Tong, E. (2015). Tasks, experiential learning, and meaning making activities: A functional approach. In M. Bygate (Ed.), *Domains and directions in the development of TBLT: A decade of plenaries from the International Conference* (pp. 157–192). Amsterdam: John Benjamins. https://doi.org/10.1075/tblt.8.06moh

Newman, D., Griffin, P., & Cole, M. (1984). Social constraints in laboratory and classroom tasks. In B. Rogoff & J. Lave (Eds.), *Everyday cognition: Development in social context* (pp. 172–193). Cambridge, MA: Harvard University Press.

Nitta, R., & Baba, K. (2018). Understanding benefits of repetition from a complex dynamic systems perspective: The case of a writing task. In M. Bygate (Ed.), *Language learning through task repetition* (pp. 295–327). Amsterdam: John Benjamins.

Ochs, E. (1988). *Culture and language development: Language acquisition and language socialization in a Samoan village*. Cambridge: Cambridge University Press.

Ochs, E. (1996). Linguistic resources for socializing humanity. In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 407–437). Cambridge: Cambridge University Press.

O'Connor, M. C., & Michaels, S. (1996). Shifting participant frameworks: Orchestrating thinking practices in group discussion. In D. Hicks (Ed.), *Discourse, learning, and schooling* (pp. 63–103). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511720390.003

Pinter, A. (2007). Some benefits of peer-peer interaction: 10-year-old children practicing with a communication task. *Language Teaching Research*, 11, 189–207. https://doi.org/10.1177/1362168807074604

Putney, L. G., Green, J., Dixon, C., Durán, R., & Yaeger, B. (2000). Consequential progressions: Exploring collective-individual development in a bilingual classroom. In C. D. Lee & P. Smagorinsky (Eds.), *Vygotskian perspectives on literacy research: Constructing meaning through collaborative inquiry* (pp. 86–126). Cambridge: Cambridge University Press.

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. New York, NY: Palgrave.

Saville-Troike, M. (2003). *The ethnography of communication: An introduction* (3rd ed.). Malden, MA: Blackwell. https://doi.org/10.1002/9780470758373

Shintani, N. (2012). Repeating input-based tasks with young beginner learners. *RELC Journal*, 43(1), 39–51. https://doi.org/10.1177/0033688212439322

Shintani, N. (2018). Mediating input-based tasks for beginner learners through task repetition: A sociocultural perspective. In M. Bygate (Ed.), *Language learning through task repetition* (pp. 269–292). Amsterdam: John Benjamins.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principles and practice in applied linguistics: Studies in honor of Henry Widdowson* (pp. 125–170). Oxford: Oxford University Press.

Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97–114). Oxford: Oxford University Press.

Swain, M. (2006). Languaging, agency and collaboration in advanced second language proficiency. In H. Byrnes (Ed.), *Advanced second language learning: The contribution of Halliday and Vygotsky* (pp. 95–108). New York, NY: Continuum.

Ushioda, E. (2008). Motivation and good language learners. In C. Griffith (Ed.), *Lessons from good language learners* (pp. 19–34). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511497667.004

van Lier, L. (1988). *The classroom and the language learner: Ethnography and second-language classroom research*. Harlow: Longman.

van Lier, L. (1996). *Interaction in the language curriculum: Awareness, autonomy, and authenticity*. Harlow: Longman.
van Lier, L. (2000). From input to affordance: Social-interactive learning from an ecological perspective. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 245–259). Oxford: Oxford University Press.
van Lier, L. (2004). *The ecology and semiotics of language learning: A sociocultural perspective*. Dordrecht: Kluwer. https://doi.org/10.1007/1-4020-7912-5
van Lier, L. (2008). The ecology of language learning and sociocultural theory. In A. Creese, P. Martin, & N. Hornberger (Eds.), *Encyclopedia of language and education, Vol. 9: Ecology of language* (pp. 53–65). Dordrecht: Springer. https://doi.org/10.1007/978-0-387-30424-3_221
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.). Cambridge, MA: Harvard University Press.
Walsh, S. (2011). *Exploring classroom discourse: Language in action*. London: Routledge.
Wells, G. (1999). *Dialogic inquiry: Towards a sociocultural practice and theory of education*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511605895
Wink, J., & Putney, L. (2002). *A vision of Vygotsky*. Boston, MA: Allyn & Bacon.
Wortham, S., & Reyes, A. (2015). *Discourse analysis beyond the speech event*. New York, NY: Routledge.

Appendix

Transcription conventions

=	latched utterances
[	beginning of overlap
(words)	words not clearly heard, (x), an unclear word
((comments))	relevant details pertaining to interaction
italics	approximate translation
underlining	emphasis or stress
:	unusually lengthened sound
.	terminal falling intonation
,	rising, continuing intonation
?	high rising intonation
–	(unattached) brief, untimed pause (i.e., less than 0.5 seconds)
(number)	a timed pause
x-	(attached on one side) cutoff often accompanied by a glottal stop (e.g., a self-correction)
"utterances/sentences"	direct or indirect quotes
boldfaced	tracers or focal utterance of point of discussion for analytical purposes

(adapted from Duff, 2002)

CHAPTER 10

Mediating input-based tasks for beginner learners through task repetition

A sociocultural perspective

Natsuko Shintani
Kobe Gakuin University

This chapter examines task repetition from a sociocultural perspective. I will report a study examining how the dynamics of teacher-student interaction change when the same task is repeated over time. I draw on sociocultural theory by viewing task repetition as one way of dynamically mediating learners' participation in an input-based task. The task was a listen-and-do task consisting of a simple procedure – the teacher's commands followed by students' action – with no requirement that the learners speak in the L2 (English). The same task was repeated nine times over five weeks. The participants were a group of six children, aged six, who had no L2 knowledge or experience of formal language learning. The lessons were audio- and video-recorded. The transcribed interactions and the video-data were analysed in detail to explore how the teacher's scaffolding changed over time and how the students shifted from other-regulation to self-regulation. The analysis showed that the teacher employed two different approaches: mediating the task commands in the early lessons to assist the students' comprehension and then pushing learners towards self-regulation in later lessons by reducing the level of support. The teacher's actions led to co-construction and enabled the students to move from other-regulation to self-regulation in both comprehending key items in the input and voluntarily producing them. Overall, the study shows how the nature of 'activity' changed as the task-as-workplan was repeated over time.

Introduction

When a task is repeated, learners may perform it in a different way. Cognitive psycholinguists explain this in terms of the learners' experience stored in their memory system. The knowledge learners develop when they encounter problems in their first performance of a task may contribute to different responses in the repeated tasks. Sociocultural theory (SCT) sees changes in the performance of a task in terms of the

DOI 10.1075/tblt.11.10shi

development of psychological tools through co-construction and scaffolding (John-Steiner, 2007). This perspective provides an alternative way of accounting for changes when a task is repeated.

The study reported in this chapter involved an unusual type of task repetition. It investigated young learners with no experience of learning a second language with consequent limits on the resources that the teacher could use. A simple input-based task was extensively repeated (nine times) with more or less the same format. The task formed the basis for nine whole lessons, involving participation on the part of both teacher and students. The study, therefore, provides an unusual perspective on task repetition, especially given that to date this has been investigated in terms of production-based tasks almost exclusively.

I will start by examining how SCT views language development. I will then consider Activity Theory, which will serve as the framework for investigating task repetition in the classroom.

Language development in sociocultural theory

Regulation and internalization

SCT sees language development as both a social and psychological concept whereby language is *developed* through social interaction and then *internalised* as a psychological tool (John-Steiner, 2007). In SCT, language is a symbolic artefact which has two functions: *interpersonal (or social) mediation* and *intrapersonal (or private) mediation*. The child's first language (L1) develops initially through interpersonal language use – most often through interaction between the child and a caretaker. Later, the child starts to use the language intrapersonally by addressing speech to the self in order to direct attention to a specific task in a focused manner. According to SCT, interpersonal activity serves as the foundation for intrapersonal functioning and internalization of new functions. Lantolf and Thorne (2007) suggest that this process of language development (i.e., internalization) involves 'imitation'. The child's imitation of a caretaker's utterance often occurs immediately after the utterance but can also occur after a delay of a few days (Tomasello, Savage-Rumbaugh, & Kruger, 1993). According to Lantolf and Thorne (2007), deferred imitation permits children to analyse language 'off-line' and points to a continuum between imitation and spontaneous language production.

Imitation occurs as either *private speech*, which is overt and observable, or *inner speech*, which is covert and hidden. Private speech is a product of a child's experience of social speech and a precursor of the 'inner speech' that informs their individual cognition. Child language development thus follows three steps: (a) Children use the language to interact with a caretaker; (b) Around the age of three or four, they start

to talk to themselves (private speech) while they play, and (c) eventually this private speech is internalized and they no longer need to verbalise out loud when thinking. The use of private speech is also observed in the process of adult L2 learning (McCafferty, 1994).

Children use private speech in their L1 often when they face difficulties or engage in cognitively demanding activities such as problem-solving tasks (Winsler, De León, Wallace, Carlton, & Willson-Quayle, 2003). Vygotsky (1987), observing private speech that appeared while children planned activities and strategies, considered private speech as a sign of using language for self-regulation of behaviour and thus indicative of development. He also noted that private speech facilitates cognitive processes, such as solving task problems, enhancing imagination, thinking and conscious awareness. A number of studies have shown that adult L2 learners also use private speech to make phenomena salient to the self (DiCamilla & Antón, 2004; Ohta, 2001). Private speech is, then, a sign of language development – it indicates a move from inter-personal mediation to intra-personal mediation.

The shift from inter-personal to intra-personal mediation can be explained by reference to *regulation* and *internalization*. Regulation refers to children's capacity to regulate their own activity through using tools (i.e., mediation). The process of developing self-regulation moves through three general stages. The first stage is *object-regulation*, when children are often controlled by or are using objects in their environment in order to think. The second stage is *other-regulation*, which involves various levels of assistance by others. The final stage, *self-regulation*, refers to the ability to complete activities with minimal or no external support. This process of shifting from the stages regulated by external assistance (object- or other-regulation) to a stage where internal resources are available to complete an activity (self-regulation) is called internalization (Donato, 1994). SCT sees language development arising as a person moves from other-regulation (i.e., language use is mediated inter-personally) to self-regulation (i.e., language use is mediated intra-personally).

SCT, then, sees internalization as taking place through the process of negotiation between individuals and their social environment. From this perspective, the goal of task-based language teaching (TBLT) is to facilitate the shift from the stage where learners need an expert's support to comprehend or produce using the L2 to the stage where they can make use of their linguistic repertoire independently to perform the task. The study reported below shows that such a shift took place when young beginner learners repeated a task nine times.

Zone of proximal development (ZPD)

The Zone of Proximal Development (ZPD) is the metaphoric notion of the site where the shift from other-regulation to self-regulation (i.e., internalisation) takes place.

Vygotsky (1987) claimed that mental abilities in life start in a stage where "the process is distributed between the individual and some other person(s) and/or cultural artefacts" and then move towards the intra-mental stage where "the capacity is carried out by the individual acting via psychological mediation" (p. 17). The notion of ZPD has been used to explain L2 learning. Swain (2000) argued that L2 development, like L1 development, arises in the process of building dialogue between individuals (interpersonally) as they collaborate in ZPD activity. One way in which this occurs is in students' imitation of teacher utterances. However, in SCT, it is not just a matter of copying an expert's skills; rather, learners "transform what the experts offer them as they appropriate it" (Lantolf, 2000, p. 17). Thus, imitation should arise only when learners voluntarily imitate the teacher's (or a peer's) utterances in the process of constructing collaborative dialogue (i.e., a meaningful conversation). Lantolf (2000) distinguished 'imitation', which he saw as contributing to development, from the mechanical repetition of teachers' utterances that can occur in traditional language classrooms. This perspective may cast an important light on the concept of task repetition.

In an often-cited study, Aljaafreh and Lantolf (1994) explored how the ZPD and mediation can be applied to L2 learning. The study identified how a tutor collaboratively created ZPDs with a student when providing corrective feedback on the students' errors in a written composition during a tutorial session. They observed, for example, that the teacher first provided implicit hints to elicit the student's self-correction and only if necessary moved towards more explicit correction, eventually withdrawing all support once the student demonstrated the ability to self-correct independently. Aljaafreh and Lantolf developed a regulatory scale to reflect the degree of explicitness in the teacher's help. They argued that for mediation to work effectively, language learning teachers need to gradually withdraw help in order to tailor it to the learners' level of development. They also argued that mediation needs to be contingent. When help is provided, its qualities depend on current needs, meaning that the explicitness of the assistance may increase over the short term in some cases, decrease or not even be necessary in other cases, while on another occasion very explicit help is still needed. In other words, by ensuring that the level of support they provide is graduated and contingent, learners can be helped to make the shift from other-regulation to self-regulation (Lantolf & Poehner, 2014; van Compernolle, 2015). Shintani and Ellis (2014) examined how such a shift took place by using 'learning-behaviour tracking' methodology (Markee, 2008). They documented how young, beginner learners repeating the same input-based tasks moved from (1) comprehending two adjectives, *big* and *small*, to (2) repeating the words after the teacher, and then finally to (3) producing them themselves. The current study re-visits the data analysed by Shintani and Ellis using the concepts of activity theory to explore the co-construction and scaffolding taking place through repeated input-based tasks.

Activity theory

Activity theory developed out of Vygotsky's work. It views human activities as systemic, purposeful interactions of the person with the world. A.N. Leontiev (1978) proposed three hierarchical levels that any activity in human behaviour involves: (a) Any action starts from a 'motive' (why do something), (b) the motive leads to some kind of action that is goal-oriented (what things are being done), and (c) the action occurs under conditions where appropriate mediational means are available (mechanics of action). The three dimensions comprising this 'activity system' (Lantolf & Thorne, 2006) are the subject of the action (i.e., the individual person), the object of the activity, and mediational means (i.e., the symbolic or material artefacts). According to A.A. Leontiev (1981), 'motives' can be biologically determined (e.g., the need to survive) or socially constructed (e.g., the need to win a game). These three levels are interconnected and thus cannot be separated. It follows that any action in an activity is intentional and meaningful and, also, the same observable activity can be linked to different goals and motives.

An example of this can be found in Wertsch, Minick and Arms (1984), who compared the interactional activities of two groups of dyads: rural Brazilian mothers and their children and urban school teachers and their students. Both groups completed a puzzle-copying task in which they were required to copy a scene depicted in a model. Wertsch et al. showed that the teacher encouraged the children to carry out the actions by themselves by instructing them on how to work with models. In contrast, the rural mothers took control over achieving the goals with the result that children rarely made errors. Wertsch et al. explained that the same *conditions* (the task and the available tools) resulted in markedly different activities because the motivations of the teachers and the mothers were different. While the teachers treated the activity as an opportunity to educate the students, the mothers' goal was to produce an error-free copy of the model.

Drawing on this theory, Lantolf and Thorne (2007) argued that "students with different motives often have different goals as the object of their actions despite the intention of the teacher" (p. 12). Even if students in the same class engage in the same task, they may not be engaged in the same activity. In this way, when a task is repeated with the same students, it is likely to be interpreted and reshaped in actual performance (Coughlan & Duff, 1994). Responding to the students' performance, the teacher might also adjust the implementation of the task. Exploring such dynamic changes will shed lights on the process feature of task repetition.

In this chapter, I will investigate whether the same tasks result in different activities when a teacher implements input-based tasks for young, beginner learners. To this end, I will examine the interactions between the teacher and the students that occurred in the nine repeated input-based tasks, and what mediational means were

used in accomplishing the action. Using the analysed data, I will also examine how the teacher and the students collaboratively constructed ZPDs to tailor the tasks to the students' level of L2 development in the repeated tasks.

The study

The current study reanalysed the data obtained for an experimental study in Japan that I reported in Shintani (2016). The original study compared the effectiveness of task-based language teaching and traditional production-based instruction. It took place in a private language school in Japan, where I had taught English to children for ten years and was motivated by my dissatisfaction with the traditional production-based approach for teaching my students. TBLT constituted an innovation in this instructional context. One of my goals of the original study was to validate the use of TBLT in such contexts.

In the project, input-based tasks were repeated nine times over a five-week period. Language learning was measured by tests assessing the students' knowledge of 24 nouns embedded in the task, 12 adjectives that occurred incidentally as the task was performed and acquisition of plural *-s*. The results showed that the participants, who were six-year-old beginner learners of English, learned to produce the vocabulary items and also to comprehend the meaning of plural *-s*. What particularly interested me was the fact that even though the tasks did not require language production, the learners still actively engaged in L2 production to request clarification when they did not understand the teacher's utterances. Gradually, their L2 production started involving a wider range of words as the task was repeated. The purpose of the current study is to examine how the teacher mediated the students' comprehension of the input provided by the tasks and their L2 production, and how this mediation evolved over the nine repetitions of the same task. The following research questions guided the current study:

RQ1: Did repeating the same task result in different actions by the teacher?

RQ2: How did the teacher and the students collaboratively construct ZPDs in the repeated tasks?

The task employed in the project was a listen-and-do task (a simple input-based task). The same task was repeated nine times over five weeks with the young beginner learners. The interactions were analysed in detail to examine how the teacher's scaffolding changed over time and the how the students shifted from other-regulation to self-regulation. Drawing on the framework of Activity Theory, I examined how the teacher's motive changed leading to different actions in the repeated task and how this

influenced the students' performance. In other words, I will show how the same *task* resulted in different *activity* (Coughlan & Duff, 1994).

Participants

In the research project, I, as the researcher, took the role of teacher. I had had experience teaching English in Japan for about ten years but had never taught the students participating in this project and had never previously used TBLT.

Six Japanese children aged six with almost no experience of second language learning participated in this project. They were recruited for this study by a small private language school owned by the researcher. This school provided the children with an opportunity to start learning English before it was introduced as a subject in primary school in grade three. The students came to the class twice a week during the project and received nine 45-minute lessons over five weeks in which they completed the same input-based tasks. As English is not usually used in daily discourse in Japan, the only exposure the children had to English was during the research project.

Teaching materials and procedures

The task for this study was input-based; that is, it required learners to listen to input and demonstrate their understanding non-verbally. However, production was not prohibited and, as we will see, the young learners engaged quite actively. The task involved the participants' listening to the teacher's commands and demonstrating their understanding by selecting the correct cards from a set of picture cards in front of them. Embedded in the task were twenty-four nouns, six of which appeared in both singular and plural forms (Table 1). For example, the students had one card with one squirrel and one card with two squirrels. The students were required to distinguish the plural form (i.e., 'squirrels') from the singular form (i.e., 'squirrel') in order to respond correctly to the commands. The target forms were the 24 nouns and plural *-s*.

Table 1. Target nouns (which appeared in the flash cards in the task)

Category	Nouns appeared in the commands
Animals (8 items + 2 plural items)	*camel, ostrich, crocodile, hippopotamus, squirrel, seal, polar bear, peacock, crocodile**s**, peacock**s***
Home appliances (8 items + 2 plural items)	*pan, ladle, chopsticks, cutting board, plate, soap, battery, toothbrush, batterie**s**, toothbrushe**s***
Fruit and vegetables (8 items + 2 plural items)	*green pepper, eggplant, chestnut, radish, leek, pear, mandarin, persimmon, pear**s**, mandarin**s***

A set of 30 picture cards representing 24 singular nouns plus six plural nouns were given to each student. The students were asked to choose the card that corresponded to the teacher's L2 command such as "Please take the crocodile (or crocodiles) to the zoo". The teacher allowed time for the students to choose their own card before asking the whole class to show the cards they had chosen. Then the teacher provided feedback on the students' choices by showing them the correct card. Those students who had chosen the correct card placed it in their individual folders. Those who had chosen a wrong card were told to put the correct card into their individual 'rubbish box'. At the end of the task, the student who had the most cards in his/her folder was the winner. At the beginning of each lesson, the goal and task procedures were explained to the participants in Japanese, but the teacher made every effort to use only English during the performance of the tasks.

The same task was conducted nine times over five weeks, twice a week. Completing the task took about 20 minutes.

Data analysis

The nine performances of the task were audio- and video-recorded. An audio-recorder was set in the centre of the table around which the six participants sat. A video-recorder was set at the back of the classroom to capture all six students. The audio data were transcribed and the video data were used to identify individual participants' utterances and non-verbal features. The transcribed data were first categorized into two general types of teacher's actions: those providing support to the students and those withdrawing support. Based on these two types of actions, the teacher's motivation and goals were determined by means of a detailed examination of the transcripts. In addition, utterances containing the teacher's and the students' use of L1 (Japanese) and L2 (English) were counted. The teacher's L1 speech was categorized based on three functions: the explanation of the task procedures, the meaning of the L2 words and various others. The students' L2 speech was first classified as either private speech (i.e., speech that was not intended to communicate with others) or social speech (i.e., speech intended to communicate with others such as asking for help). Ohta (2001) proposed three defining characteristics of private speech: (1) the volume was reduced, (2) the speech was not produced in response to a question/comment directed specifically to the individual by the teacher or another student, and (3) it did not receive a response by the teacher or classmate (p. 38). In the current study, however, students sometimes spoke to themselves in the same volume as they speak to others. Also some researchers argue that private speech might receive a response when it matches the teacher's intention at a particular moment (Ishii, 1997). In this analysis, therefore, only the second criterion of Ohta (2001) was used to identify private speech – that is, speech consisting of audible self-directed utterances where the speaker has no

intention of communicating with another individual. The other type – the students' social speech – was further categorized into four types (those using the word *colour*, those using colour adjectives, those using numerical words, and other words). The researcher completed all the coding and another coder separately coded 10% of the transcripts. Any disagreements were resolved through discussion.

Findings

Activity Theory suggests that any action starts from some kind of motive which then leads to some kind of goal-directed action. The analysis showed that in the early lessons, the teacher focused on mediating the learners' completion of the task, but in later lessons, the teacher's focus shifted to pushing the students to self-regulate in performing the tasks. I will examine these two strategies separately.

Mediating the learners' completion of the task

In the first few lessons, the teacher's actions seemed to exclusively focus on helping the students complete the task smoothly and successfully. This was achieved by providing means of various mediational devices whenever the teacher found the students having problems comprehending the task commands. Excerpt 1 shows the interaction resulting from the very first task command in Lesson 1. It shows that the six-year-old beginner learners had difficulty identifying the correct card. The students responded to the teacher's command in Japanese (lines 2 and 4), indicating that they did not comprehend. The teacher then employed a variety of mediational devices to help the students understand.

The teacher first used an elaboration of the target noun 'peacock' by providing information about the colour of the item ("the peacock is blue" in line 6). Recognizing that this was not successful, the teacher used another mediational device. She provided the L1 translation of the words 'listen' and 'hint' to call for attention and repeated the word 'blue' (line 8). S2 seemed to utilise this clue and chose the correct card (line 9). However, some students were still unable to understand. The teacher then resorted to another strategy – using ostensive definition by pointing to a blue object in the classroom (line 10). S3, however, did not respond to this clue but instead asked in Japanese if the item was an animal. When the question was repeated by another student (S2) in line 26, the teacher resorted to the L1 again to make it clear that she was referring to the zoo. She then repeated the command but ultimately in this first performance of the task, the children were not successful in understanding it.

Excerpt 1. (Lesson 1)

1. T: okay, let's try. please take the peacock to the zoo.
2. S1: *peaco? wakannai* (tr: I don't know that).
3. T: please take the peacock to the zoo.
4. S2: *wakatta* (tr: I know) (talking to S4 who was sitting next to him), *kaba, kaba* (tr: hippopotamus, hippopotamus).
5. S3: (talking to the teacher) *dobutsu*? (tr: animal)
6. T: listen listen listen. okay? the peacock is blue
7. S2: blue
8. T: …. everyone, listen. do you understand 'listen'? *kiite, tte iukoto ne* (tr: it means' listen). hint. hint. hint. peacock. the peacock is blue. a blue bird.
9. S2: *matte, yameta* (tr: wait, I'll change it) (looking at his cards on the table), *wakatta* (tr: I know), (humming and smiling while he holds the card).
10. T: do you understand blue? blue is this (pointing a blue object in the classroom). okay? blue.
11. S3: *dobutsu* (tr: animal)?
12. T: I said "the zoo" right? *"zoo" wa dobutsuen tte koto ne* (tr: "*zoo*" means zoo).
13. S2: *wakattakara* (tr: now I know, so..). *blue tte ittekurete arigatou* (tr: thank you for telling me it's blue).
14. T: blue. blue. blue. blue. okay? ready? let's check. three, two, one, go!. (showing the correct card to the class), peacock.
15. S2: peacock.
16. T: peacock is this one. blue, right?
17. S2: yeeey.
18. S5: *matchigaechatta* (tr: I got it wrong).
19. T: I said blue right?
20. S2: *naniiro* (tr: what colour)? *tsugi naniiro ka oshiete* (tr: tell me the colour of the item for the next question).
21. S4: *naniiro ka oshiete* (tr: tell me the colour).
22. T: no no no no. okay S3, S5, S6, find the peacock, find the peacock
23. S3: *naiyo, naiyo* (tr: I don't have one).
24. T: *aru aru* (tr: you do), (walking to the students seats and help her find the card), this one. this one. put them into the rubbish box (demonstrating to the class). *gomibako ni ireru* (tr: put it into the rubbish box).
25. T: okay next one. okay? next one. please take the squirrels to the zoo.

26. S3: *dobutsu* [animal]?
27. T: squirrel to the zoo.
28. S5: squirrel
29. T: yes, that's right. that's right.
30. S2: *dobutsu* [animal]?
31. T: yeah (.) I said zoo right? zoo. *dobutsuen ni iku to iukotowa* [going to the zoo means].
otosanaide [don't drop it]. okay. squirrels. squirrels. ready? three (.) two (.) one (.) go.
32. Ss: (showing their cards but no one was correct)
33. T: squirrels. squirrels. that's a crocodile. squirrels are this one. listen listen. squirrels are these ones. put them into the rubbish
box.

The students engaged in various actions – requesting help from the teacher, using the L1 and paying attention to the teacher's clues. The entire sequence of actions is focused on solving the communicative problem. Such actions were recognized by the teacher and led to the use of various mediational devices to assist task completion. I will now focus on the three types of mediation that the teacher used in this excerpt – comprehension strategies, elaboration and the use of the L1 – and examine how her use of these devices changed in the repeated tasks.

Teacher's action 1: Comprehension strategies

The above excerpt showed that the teacher employed several strategies to mediate learners' comprehension of 'zoo' in her commands. Later in Lesson 1, the teacher explained the meaning of the word 'supermarket' – the other key word for understanding the location in the commands. Initially there was no evidence that the students understood the meanings of 'zoo or supermarket'. However, as Excerpt 2 illustrates, students later showed they had picked up these words and could use them to request clarification. In Lesson 3, for example, S3 imitates the teacher's use of 'supermarket'. This example shows that the students were able to make use of the teacher's comprehension strategy to help identify the location that a specific command referred to.

Excerpt 2. (Lesson 3)

1. T: please take the pan to the supermarket.
2. S1: pan
3. S3: supermarket?
4 T: yes, to the supermarket.

Ultimately, the students were able to utilise the locational information (zoo or supermarket) more fully. In Excerpt 3, taken from the last lesson, S1 demonstrated that he

had successfully identified the referent of the teacher's command (i.e., 'crocodile') but also that he had noticed the teacher's error in referring to the supermarket rather than the zoo. This student produced "supermarketto" with a rising intonation to point out the teacher's error. In turn 4, S2, who was sitting next to S1, used his limited English to make fun of the idea of a crocodile appearing in the supermarket setting. The teacher then extended her utterance, causing both students to laugh. This indicates that by the end of the nine repeated lessons, the students not only understood the word 'supermarket' but were able to use it productively and creatively in their own utterances.

Excerpt 3. (Lesson 9)

```
1.  T:      please take the crocodile to the supermarket.
2.  S1:     (taking the 'crocodile' card from his table) supermarketto?
3.  T:      oh, sorry, zoo, zoo. not supermarket.
4.  S2:     crocodile, supermarket (gestures).
5.  T:      yeah, it's scary isn't it, if a crocodile is in the
            supermarket.
6.  S1, S2: (laugh)
```

Teacher's action 2: Elaborating a key word

The teacher provided clues to help the learners identify the key referents in her commands, for example, by stating the colour of a referent (cf. line 6 and line 8 in Excerpt 1). This way of elaborating on the noun apparently helped the learners' comprehension as indicated by S2's utterance in line 13 (*wakattakara* [tr: now I know, so.] *blue tte ittekurete arigatou* [tr: thank you for telling me it's blue]). The teacher used a variety of adjectives to elaborate on nouns. In Lesson 1 alone, she made use of the following: *brown* (13 times), *blue* (10), *white* (5), *black* (4), *purple* (4), *big* (3), *grey* (2), *heavy* (1), *fruit* (1) and *orange* (1). Most of these adjectives were colour words, which seemed to help the students understand the commands.

The students appropriated the teacher's elaboration strategy by asking the teacher (in Japanese) to tell them the colour of the key referents. This occurs in the very first excerpt (see lines 20 and 21). In Lesson 2 (see Excerpt 4), one student explicitly uses Japanese to ask the teacher to tell him the colour of the referent. When the teacher repeats "colour", the student responds by imitating this word, leading the teacher to name the colour of the referent. In Excerpt 5, which occurred approximately 90 seconds after Excerpt 2, the same student (S1) asks the colour of an item in English (line 2). Another student repeats the utterance (line 3) and the teacher responds by providing the information (line 4). Subsequent to this interaction, the students produced the English word "colour" 12 times in total to request clarification. We can see evidence of co-construction and scaffolding here then: evidence of the learners actively encourag-

ing the teacher to elaborate the nouns and also of the important role that imitation can play in assisting internalisation.

Excerpt 4.

1. T: okay the next. please take the mandarins to the supermarket.
2. S1: *iro wa* (What is the colour?)
3. T: the colour? The colour?
4. S1: nodding) colour
5. T: mandarins are orange.

Excerpt 5.

1. T: okay the next. please take a chestnut to the supermarket.
2. S1: colour wa? (tr: what 'colour' is it?)
3. S2: colour, colour
4. T: brown. do you remember? brown. (pointing an object). this colour.

Another type of elaboration involved plural nouns. As noted in the method section, one of the linguistic goals of the task was to introduce plural forms. The teacher used an elaboration strategy similar to the one described above to help the students distinguish singular and plural forms, as shown in Excerpt 6 from Lesson 1. In this case, the teacher provided the number word *three*. Again, we find the students appropriating this strategy by using number words themselves to request clarification, as in Excerpt 7 from Lesson 2. Overall in Lesson 2, students uttered numbers (i.e., *one*, *two*, *three* and *four*) nine times to help them identify whether the referent was singular or plural, with the teacher responding by confirming the number. The use of this strategy by the students became more frequent over time – Lesson 3 (20 times), Lesson 4 (28 times) and Lesson 5 (32 times). Thus, a strategy initially employed by the teacher was seized upon by the students as they became increasingly active in seeking the clues they needed to comprehend the commands.

Excerpt 6.

1. T: please take that crocodiles to the zoo. please take that crocodiles in the zoo. crocodiles. crocodiles.
2. S3: crocodile
3. T: crocodile<u>s</u>, so there are three (indicating 'three' with her fingers). three animals in one card. okay? three, two, one, go!
4. Ss: (showing their cars)
5. T: S2, too late. crocodiles. (checking students' answers) you are correct. i said three right? find the crocodiles.

Excerpt 7.

1. T: please take the crocodile to the zoo. crocodiles are very big heavy scary animal.
2. S3: one?
3. T: one. that's right. crocodile.

Teacher's action 3: L1 use

As shown in Excerpt 1, the teacher's use of the L1 played an important mediating role in Lesson 1, serving a variety of functions. Sometimes the teacher provided the Japanese translation of a key word (e.g., *zoo*) or of a whole utterance (do you understand 'listen'? *kiite, tte iukoto ne* tr: it means listen). The teacher also used the L1 to explain procedures as in line 24 (tr: put it into the rubbish box). As might be expected, the teacher used Japanese quite frequently to begin with, but over time this decreased. Figure 1 shows the teacher's use of the L1 in Lessons 1, 3, 5, 7 and 9, for (a) task procedures, (b) translating L2 words, and (c) other uses such as responding to students' personal comments (e.g., Student: *kyou kore tabeta* [I had this today]; Teacher: *kyou, honto* [today? really] in Lesson 1). The reduction in the teacher's use of Japanese occurred in parallel to the students' growing ability to appropriate comprehension strategies to understand the teacher's commands.

In sum, the above analysis shows how the teacher's actions (in the sequence of strategic moves realised through different types of utterances) helped students overcome their problems in completing the task and how the students appropriated the strategies the teacher used to help them. In this way, the students became increasingly responsible for their own comprehension, which removed the need for the teacher to fall back on Japanese. In the next section, I will point out that as the task was repeated,

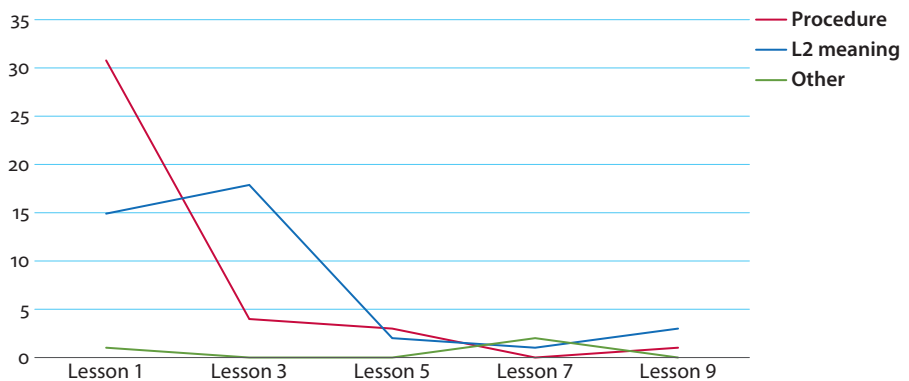

Figure 1. The frequency of the teacher's L1 utterances

the nature of the activity changed as the teacher's motive switched from mediating comprehension to pushing the learners' own production of the L2.

Pushing learners to self-regulate

In this section, three sets of teacher actions will be examined: encouraging students to request clarifications using adjectives, pushing students to comprehend the plural morpheme, and encouraging the students to use the L2. I will examine each of them in detail.

Encouraging students to request clarifications using adjectives

As shown in Excerpt 5, the students started to use the L2 word 'colour' to negotiate for meaning in Lesson 2. This negotiation strategy continued in Lesson 3. However, at some point in Lesson 3, the teacher refused to answer to the students' clarification requests using the word 'colour', as shown in Excerpt 8. In line 3, the teacher indicated that she would not answer the student's question and demonstrated that the students could ask a question by using adjectives. As her explanation in the L2 was misunderstood by S5 (line 4), she added an explanation in Japanese (line 5). This was understood by the students, who then started to request clarification using L2 adjectives.

Excerpt 8. (Lesson 3)

1. T: okay the next. please take the crocodile to the zoo. crocodile to the zoo. crocodile.
2. S2: colour?
3. T: uh, no hint. too easy. no hint, no hint. you can ask me questions, 'blue'?
4. S5: blue (looking at the cards)
5. T: no no no no. you (emphases) can ask me questions. *Situmon shite iiyo* [tr: you can ask me a question]. blue?
6. S2: blue?
7: T: no.
8: S4: black?
9: T: no
10: S2 green?
11: T: yes! green.

At this point, the teacher's motive seems to have shifted from task completion to pushing learners to self-regulate by actively negotiating for meaning in English. More specifically, the teacher's goal became that of equipping students with the various English adjectives for negotiating for meaning in the task.

After this sequence, the students stopped using 'colour?' to request clarification and started to use colour adjectives instead. The frequency of students' clarification requests using colour adjectives increased as the task was repeated: there were 8 times in Lesson 2, 22 in Lesson 3, 32 in Lesson 4, 15 in Lesson 6, 31 in Lesson 7, 38 in Lesson 8 and 60 in Lesson 9. The adjectives uttered by the students included 'black' (41 tokens in total), 'blue' (37 tokens), 'brown' (44 tokens), 'grey' (24 tokens), 'purple' (15 tokens) and 'white' (45 tokens).

Teacher's action 2: Pushing students to comprehend the plural morpheme

The analysis of Excerpt 1 also shows that the teacher frequently mentioned a number when the command required the students to distinguish the singular and plural forms of nouns (e.g., students had a card with one squirrel and a card with three squirrels). As we have seen, such support led to the students' frequent production of these numbers to clarify the targeted item. In early lessons, the teacher provided answers when requested, which often enabled students to identify the correct cards. In Lesson 5, however, the teacher stopped responding to the students' clarification requests using numbers such as "one?" or "two?". As in Excerpt 9, she refused to answer the students' questions (lines 9 and 11) and instead explicitly contrasted the singular and plural forms of the noun (line 11). However, only S2 utilised this explicit help to identify the correct card.

Excerpt 9. (Lesson 5)

1. T: okay listen listen. please take the squirrels to the zoo
2. S2, S5: … squirrels
3. T: squirrels to the zoo. squirrels, squirrels..squirrels are kind of brown.
4. S2: … brown.
5. T: … brown yes.
6. S5: (showing 'small' with his hands)
7. T: … small yes. Small and brown. Okay.
8. S2: … one?
9. T: listen.
10. S1: … one?
11. T: listen. (showing the singular and the plural picture cards in turn) squirrel, squirrels. squirrel, squirrels. okay? please take the squirrels to the zoo. okay? okay? three (.) two (.) one (.) go (2.0)
12. Ss: (showing their own cards but only S2's card was correct)
13. T: (checking the students' choices) you don't understand (.) only S2.

It seems that the teacher's goal in the early lessons was to enable the students to complete the task. To this end she simply provided them with the numerical information they needed to complete the task, but in so doing she removed the need for them to attend to the plural morpheme. However, in Lesson five she changed her strategy, drawing the learners' attention to the difference between the singular and plural nouns forms very explicitly when she saw that the students were unable to comprehend her command after she had withdrawn the numerical information. It is worth noting, however, that the teacher still did not resort to any metalingual explanation, presumably because she deemed this unsuitable for such young learners.

Following Lesson 5, the teacher continued to ignore students' requests for help when they used a number word and again offered an explicit presentation of the singular and plural nouns in conjunction with displaying the relevant cards. Such demonstrations occurred eight times in total (three times in Lesson 5, twice in Lesson 6, twice in Lesson 7 and once in Lesson 8). Excerpt 10 presents one of these occasions. Responding to the teacher's command, S4 asked whether the card include two items (line 2). Assuming that the teacher would then demonstrate the difference between the singular and plural noun, S3 tried to stop her (line 3), presumably because he did not want S4 to notice the plural form and thereby complete the task successfully. However, the teacher ignored S3 and went ahead with the explicit demonstration of the plural form (line 4). This led to S4 noticing the morpheme (line 5) and pronouncing the /z/ morpheme. S3 then showed his disappointment. We can see from this excerpt that some students noticed and understood the morphological form more quickly than other students. Reflecting the fact that plural *-s* constitutes a difficult grammatical form for Japanese learners (Athanasopoulos & Kasai, 2008; Hakuta, 1978), some students needed the demonstration of its meaning repeatedly before they could notice the form and utilise the information in comprehending a command.

Excerpt 10. (Lesson 7)

1. T: next. please take the squirrels to the zoo. squirrels to the zoo.
2. S4: two?
3. S3: ichadame [don't tell them].
4. T: hint hint. (showing the singular and plural cards in turn) one squirrel, two squirrels. listen listen. one squirrel, two squirrels (1.0). please take the squirrels to the zoo.
5. S4: /z/
6. S3: itchadame tte ittanoni [I said don't tell them]
7. T: please take the squirrels to the zoo (.) three (.) two (.) one (.) go.
8. Ss: (showing their cards. Five out of nine students were correct)
9. T: that's right.

In Lesson 9, the teacher indicated that she was going to completely withdraw support for processing the plural form. In Excerpt 11, ignoring S2's request, the teacher closed the sequence by simply repeating the command and requesting the students to show their selected cards. Only after showing them the correct card did she use a number word to show the noun was plural and only after repeating the plural noun by itself three times. This excerpt also indicates that some students continued to be unable to distinguish plural from singular nouns even after the teacher's repeated demonstrations of the plural form.

Excerpt 11. (Lesson 9)

1. T: okay the next. no hint. please take the squirrels to the zoo.
2. S3: … two squirrels?
3. S5: … brown?
4. T: yeah, brown (.) but, about the numbers (.) no hint (.) no hint (.) I'm not going to give you any hint.
5. S2: … one, two?
6. T: please take the squirrels to the zoo. squirrels. squirrels. okay? three (.) two (.) one (.) go (1.0).
7. Ss: (showing their cards. S2 and S5 were wrong but the others were correct)
8. T: (showing the correct card) squirrels (.) squirrels, squirrels (.) two squirrels.

I have noted previously that the students' motive for completing the task was to be the 'winner' by selecting the most correct cards. However, in Excerpt 12 in Lesson 9 we can see that for some of the children, other motives were in play. A student would sometimes help another student instead of concentrating on identifying the correct card. S3 saw that the S2 could not comprehend 'toothbrushes' and so directed his attention to the plural form by saying, "it's /z/".

Excerpt 12. (Lesson 9)

1. T: okay the next. please take the toothbrushes. please take the toothbrushes to the supermarket. toothbrushes.
2. S2: nanko? nanko? (tr: how many? how many?)
3. T: listen listen. please take the toothbrushes.
4. S2: (holding the card with one toothbrush on it) ikko dayo (tr: It's one)
5. S3: (talk to S2) sanko dayo (tr: it's three)
6. S2: ikko dayo (tr: it's one)
7. S3: /z/ dayo (tr: it's /z/)
8. S2: (changes his card to the card with three toothbrushes)

Here, S2 provided scaffolding for his peer by sharing his knowledge and enabling him to achieve the task goal. These excerpts show that the construction of ZPDs was a collaborative endeavour not only between the teacher and the students but also between the students themselves.

Teacher's action 3: Encouraging the students to use the L2

In the later lessons, the teacher sometimes encouraged the students to use the L2 when they resorted to the L1 in negotiation. Excerpt 13 shows such an occasion. Responding to S3, who asked whether the item was heavy in the L1 (line 6), the teacher provided the L2 equivalent. Later, the same student (S3) asked the same question but this time used the L2 (lines 10 and 12). The teacher confirmed the student was right in line 13. In this sequence, the teacher also provided more explicit encouragement for using English. Responding to S4's clarification request in the L1 (line 16), the teacher requested S4 to use the L2. Another student (S2) produced the L2 in line 18, which was acknowledged by the teacher in line 19.

Excerpt 13. (Lesson 4)

	1.	T:	okay the next. please take the camel.
	2.	S1:	camel
	3.	T:	camel to the zoo. the camel is also.
	4.	S2:	black?
	5.	T:	not black. not black. very big.
	6.	S3:	*omoi*? (heavy?)
	7.	T:	yeah I think so. heavy and…
	8.	S4:	pata pata suru (does it flap?)
	9.	T:	*Pata pata suru*? that's Japanese. that's Japanese.
	10.	S3:	heavy?
	11.	T:	uh?
	12.	S3:	heavy?
	13.	T:	I think so, heavy. and big.
	14.	S2:	blue?
	15.	T:	no no no.
→	16.	S4:	*chairo*? (brown?)
→	17.	T:	*chairo* is Japanese. do you remember?
→	18.	S2:	brown?
→	19.	T:	brown yes. brown. brown, big, heavy. are you ready?
			camel. camel. three, two, one, go.
	20:		All students: (showing their cards)
	21.	T:	(checking the cards) okay.

As evident in Excerpt 14 from Lesson 9, by the end of the research project, the students eliminated use of the L1 as they were now able to actively negotiate the meaning of nouns using English adjectives.

Excerpt 14. (Lesson 9)

1. T: so next one. please take the leek, leek, to the supermarket. it's a long vegetable, a long one.
2. S1: brown?
3. T: brown? no.
4. S2: fruit?
5. T: fruit? no no, I said vegetable, right?
6. S6: yellow?
7. T: yellow? no.
8. S4: supermarket?
9. T: yes, it should go into the supermarket, yes.
10. S3: green.
11. T: yes green.
12. S3: white.
13. T: right. green and white. okay? leek. three, two, one, go!
14. Ss: (holding their cards up)
15. T: (showing the correct card) right, everyone is correct.

Figure 2 provides an analysis of the types of words that the students used to negotiate understanding of the commands. There were no instances in Lesson 1. In Lesson 2, the students only used the word *colour* (e.g., "colour?") and numerical words (e.g.,

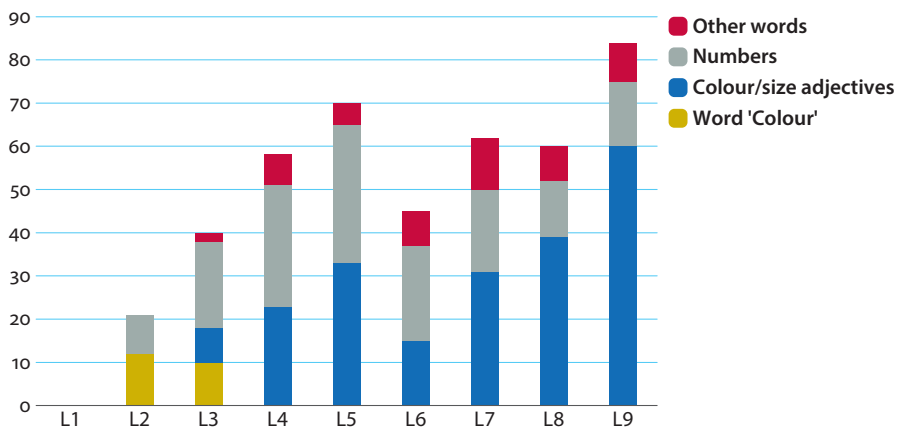

Figure 2. Students' negotiation of meaning. Key: Word 'colour': asking the colour of the object using the word 'colour'; Colour/ size adjectives: use of colour or size adjectives (e.g., "blue?", "big?"); Numbers: use of numerical words (e.g., "two?"); other words: use of other words (e.g., "supermarket?")

"two?"). The use of the word *colour* stopped in Lesson 3. Instead, the student now produced colour and size adjectives, which remained frequent right up to the last lesson. This figure also shows the increase in the students' use of negotiation as the task was repeated and they became familiar with the task procedures.

Sometimes students would use their L1, possibly as a psychological tool, to regain regulatory control and/or solicit assistance. The teacher would follow the students' lead by translating words into Japanese but subsequently encouraged them to use the L2. The students' active production suggests that the students maintained their motive to 'win the game' throughout the nine repetitions of the task. The competitive nature of the task motivated the students to engage in the task and to struggle to comprehend the teacher's command by seeking help in different ways. However, the excerpts also indicate that some students manifested a different motive.

Finally, it is clear from the eager way in which the multiple students asked questions in the L2 (Excerpt 14) and sometimes rephrased their L1 utterances into L2 (S3's "heavy" and S2' "brown" in Excerpt 13) that they enjoyed not just the task itself but working to complete it using English. Evidence for this includes the increased number and variety of interrogative utterances (e.g., "white?") over time – as shown in Figure 2.

Conclusion

The analysis has shown that in the task-based classroom, the teacher's actions are informed by socially situated motives, as suggested by Activity Theory. The teacher's actions were embedded in purposeful interaction with the students, using specific tools, which were sometimes derived from the L1 but mostly from the L2, as well as the task materials and objects in the classroom environment. One motive was to enable the students to complete the task successfully but there was also a second motive of assisting the children's language development. The second goal was achievable only by succeeding with the first. The teacher's actions followed a fairly well-defined sequence: (a) identifying the students' comprehension problems, (b) providing support to enable them to solve the problems, (c) confirming whether they had overcome the issue with the teacher's help and (d) withdrawing help to encourage self-regulation. In the early lessons, the teacher's motive was to help students complete the task successfully but when the students became able to do so, the motive shifted to a higher level – to push the students to use the L2 in order to complete the task.

The analysis also showed that the teacher's purposeful actions helped the students move from other-regulation to self-regulation. For example, there was the shift from the students needing to use the L1 to negotiate to being able to use the L2. There was also the shift from the students completing the task successfully with the help of the teacher's elaboration of the key nouns to voluntarily requesting clarification using

L2 words and, for some of the children, to being able to comprehend the commands successfully without the need for negotiation. There were multiple occasions where the teacher – and sometimes the students – provided scaffolding to enable another student to manage the task.

As shown in the excerpts, the repetition of the tasks afforded the shift from other- to self-regulation and in so doing resulted in very different kinds of 'activity'. Thus, although the 'task' remained the same, the 'activity' evident in Lesson 1 (Excerpt 1) and Lesson 9 (Excerpt 14) is clearly different. The early lessons centred around the teacher assisting the students using various mediational strategies – elaborating the target words, using the L1 and utilizing comprehension strategies. In the later lessons, the teacher focused on pushing the students to self-regulate by withdrawing elaboration, requesting the use of the L2 and explicitly directing the students' attention to the linguistic form. Thus, although the task remained constant throughout, the teacher's actions evolved throughout the iterations of the task, as did the students' responses. As a result, the task gave rise to gradually changing activities through the sequence of nine lessons, activities that differed, but that nonetheless retained something in common.

As we can see from the various excerpts, the repetition of the same task did not bore the students. The interaction in Lesson 9 suggests that the students were still actively engaged in completing the task. The excerpt from this lesson even gives the impression that the students wanted to continue to negotiate for meaning even though they had comprehended the command (i.e., understood the noun in the command). It seems that the students now enjoyed negotiating in English for its own sake – a sure sign of the self-regulation that they had achieved. They were motivated not just to complete the task but also perhaps to engage as conversationalists in English. As the task became more manageable through repetition, these students developed confidence and pleasure in using English in much the same way as what happens in the repetitive experiences that children enjoy when acquiring their L1. Repeating the same task with these children had positive cognitive and affective outcomes that can be most readily accounted for in terms of SCT and Activity Theory.

References

Aljaafreh, A. L. I., & Lantolf, J. P. (1994). Negative feedback as regulation and second language learning in the Zone of Proximal Development. *The Modern Language Journal*, 78(4), 465–483. https://doi.org/10.1111/j.1540-4781.1994.tb02064.x

Athanasopoulos, P., & Kasai, C. (2008). Language and thought in bilinguals: The case of grammatical number and nonverbal classification preferences. *Applied Psycholinguistics*, 29(1), 105–123. https://doi.org/10.1017/S0142716408080053

Coughlan, P., & Duff, P. (1994). Same task, different activities: Analysis of SLA from an activity theory perspective. In J. P. Lantolf & G. Appel (Eds.), *Vygotskian approaches to second language research* (pp. 173–194). Norwood, NJ: Ablex.

DiCamilla, F. J., & Antón, M. (2004). Private speech: a study of language for thought in the collaborative interaction of language learners. *International Journal of Applied Linguistics*, 14(1), 36–69. https://doi.org/10.1111/j.1473-4192.2004.00053.x

Donato, R. (1994). Collective Scaffolding in Second Language Learning. In J. P. Lantolf & G. Appel (Eds.), *Vygotskian approaches to second language research* (pp. 33–56). Norwood, NJ: Ablex.

Hakuta, K. (1978). A report on the development of the grammatical morphemes in a Japanese girl learning English as a second language. In E. Hatch (Ed.), *Second language acquisition* (pp. 132–147). Rowley, MA: Newbury House.

Ishii, M. (1997). Kyooshitsu danwa no fukusuu no bunmyaku [Multiple contexts of classroom discourse]. *Nihongogaku [Studies of Japanese Language]*, 16(3), 21–19.

John-Steiner, V. (2007). Vygotsky on thinking and speaking. In H. Daniels, M. Cole, & J. Wertsch (Eds.), *The Cambridge companion to Vygotsky* (pp. 136–152). Cambridge: Cambridge University Press. https://doi.org/10.1017/CCOL0521831040.006

Lantolf, J. (2000). *Sociocultural theory and second language learning*. Oxford: Oxford University Press.

Lantolf, J., & Poehner, M. (2014). *Sociocultural theory and the pedagogical imperative in L2 education*. New York, NY: Routledge.

Lantolf, J., & Thorne, S. (2006). *Sociocultural theory and the genesis of second language development*. Oxford: Oxford University Press.

Lantolf, J., & Thorne, S. (2007). Sociocultural theory and second language learning. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 201–224). Mahwah, NJ: Lawrence Erlbaum Associates.

Leontiev, A. A. (1981). *Psychology and the language learning process*. Oxford: Pergamon.

Leontiev, A. N. (1978). *Activity, consciousness, and personality*. Englewood Cliffs, NJ: Prentice Hall.

Markee, N. (2008). Toward a learning behavior tracking methodology for CA-for-SLA. *Applied Linguistics*, 29(3), 404–427. https://doi.org/10.1093/applin/amm052

McCafferty, S. (1994). Adult second language learners' use of private speech: A review of studies. *The Modern Language Journal*, 78(4), 421–436. https://doi.org/10.1111/j.1540-4781.1994.tb02060.x

Ohta, A. S. (2001). *Second language acquisition processes in the classroom: Learning Japanese*. Mahwah, NJ: Lawrence Erlbaum Associates.

Shintani, N. (2016). *Input-based tasks in foreign language instruction for young learners*. Amsterdam: John Benjamins.

Shintani, N., & Ellis, R. (2014). Tracking 'learning behaviours' in the incidental acquisition of two dimensional adjectives by Japanese beginner learners of L2 English. *Language Teaching Research*, 18(4), 521–542. https://doi.org/10.1177/1362168813519885

Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97–114). Oxford: Oxford University Press.

Tomasello, M., Savage-Rumbaugh, S., & Kruger, A. C. (1993). Imitative learning of actions on objects by children, chimpanzees, and enculturated chimpanzees. *Child Development*, 64(6), 1688–1705. https://doi.org/10.2307/1131463

van Compernolle, R. (2015). *Interaction and second language development: A Vygotskian perspective*. Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.44

Vygotsky, L. S. (1987). *The collected works of L. S. Vygotsky, Vol. 1: Problems of general psychology*. New York, NY: Plenum Press.

Wertsch, J. V., Minick, N., & Arms, F. (1984). The creation of context in joint problem-solving. In B. Rogoff & J. Lave (Eds.), *Everyday cognition: Its development in social contexts*. Cambridge, MA: Harvard University Press.

Winsler, A., De León, J. R., Wallace, B., Carlton, M. P., & Willson-Quayle, A. (2003). Private speech in preschool children: Developmental stability and change, across-task consistency, and relations with classroom behavior. *Journal of Child Language*, 30, 583–608. https://doi.org/10.1017/S0305000903005671

CHAPTER 11

Understanding benefits of repetition from a complex dynamic systems perspective

The case of a writing task

Ryo Nitta and Kyoko Baba
Nagoya Gakuin University / Kinjo Gakuin University

This classroom-based study explored the benefits of repeating a writing task over one year from the perspective of complex dynamic systems. The study specifically investigated how students' use of self-regulation related to changes in L2 writing. Data were collected from 26 students in an EFL classroom at a university over 30 weeks. Students were engaged in a 10-minute timed-writing task on a chosen topic with immediate self-reflection every week. Students' L2 compositions were analysed using fluency, syntactic and lexical measures, and their self-reflections written in L1 were analysed in terms of self-regulatory processes. The analysis of two focal students revealed that the first student showed more elaborate engagement and employed self-regulatory cycles of goal-setting and self-evaluation, which improved his L2 writing over time, while the second student with more limited engagement employed less elaborate self-regulatory processes, which reflected little change in his writing. Based on these cases, we contend that repeated encounters with tasks over extended periods create a valuable pedagogic environment, and within this context students' agentic attitudes towards the L2 writing task are likely to influence their learning in significant ways.

Introduction

This chapter explains how we see repetition as related to learning in a writing task. There are a number of grounds for thinking that repetition could be an essential element of language learning in the oral modality. Repetition by caretakers who read picture books to children, for example, provides support in helping them to familiarise themselves with their first language (L1; Bruner, 1983). Repetition of a language learning task helps in the learning of a second language (L2) in the same way by providing learners with "a recognisable and recurring thematic space to which they can return to work orally again and again" (Bygate, 2006, p. 172).

DOI 10.1075/tblt.11.11nit

The effects of repeating an oral task have been generally explained using a "computational metaphor of acquisition" (Ellis, 1997, p. 477): Learners who have processed conceptual aspects of their production on the first occasion are able to free up some of their cognitive capacity to enable their attention to focus on other aspects of the language on subsequent occasions. Based on this assumption, previous research has suggested that repetition can have positive effects on L2 oral performance, but its scope has been limited, at least in two important respects. First, because data were generally collected from only few repetitions over relatively short periods, little is known about the way in which repetition is related to longer term learning, especially when a particular type of task is repeatedly provided in a classroom over an extended period. Second, because previous research on task repetition has mainly focused on the oral domain, it is less clear whether and to what extent these findings apply to the case of writing.

To address these gaps, we conducted a series of studies that investigated how learners' writing changed through task repetition over one year using the perspective of complex dynamic systems (Baba & Nitta, 2010, 2012, 2014; Nitta & Baba, 2014, 2015). We found that, while writing task repetition did not necessarily improve their command of language immediately, over time it was likely to encourage their learning in terms of fluency (Baba & Nitta, 2014) and lexical and grammatical complexity (Nitta & Baba, 2014). While we acknowledged that repetition of tasks influenced their learning in different ways from oral tasks, we also observed that learners' agentic attitudes towards tasks might also play a significant role.

We analysed our data set that was collected from a large study on the nature of learner agency to build on it and extend our previous findings, and here we discuss how repetition of a writing task can provide *affordances* for different ways of learning by different learners over time. We first review previous research on oral task repetition. Next, we explain how research on writing task repetition, from the perspective of complex dynamic systems, constitutes an alternative approach to describing/investigating the relationship between repetition and learning. We then present our findings and claim that repeated encounters with a writing task create a valuable pedagogic environment in which learners can learn in their own ways over time.

Previous research on oral task repetition

As this volume shows, much research on task repetition has been carried out as a useful pedagogic task implementation, which can upgrade learners' performance. Drawing on Levelt's (1989) speaking model, Bygate (2001) explained how task repetition can facilitate L2 oral performance: The repeated experience (i.e., exact repetition) of carrying out speaking processes (e.g., conceptualisation, formulation, and

articulation) on a particular topic may lead to freeing up some of the learners' capacities on subsequent occasions, enabling them to focus attention on other aspects of the task, particularly in the processes of formulation and articulation. Prior experience of handling discourse types in a particular task frame but with different content (i.e., procedural repetition) may be stored in long-term memory, potentially contributing to improving some aspects of performance.

This psycholinguistic orientation has produced numerous studies which have yielded robust findings on the benefits of exact repetition. The benefits found for procedural repetition, on the other hand, have been rather mixed and inconclusive. Bygate (2001) and Gass, Mackey, Alvarez-Torres and Fernández-García (1999) investigated the effects of repeating monologic tasks and found clear linguistic gains when learners were asked to recount exactly the same content, but they also discovered that these benefits did not carry over to the performance of the same type of tasks with different content (i.e., procedural repetition). Although the effects of procedural repetition seem to be limited in terms of complexity, accuracy and fluency (CAF) measures, there is some evidence of its benefits on other dimensions of interactional tasks. For example, Mackey, Kanganas and Oliver (2007) found that procedural repetition provided young learners with more opportunities to use feedback. Kim (2013) also found that adolescent learners in a procedural repetition group were more engaged in collaborative talk about the language than those in an exact repetition group (see Table 1 for a summary).

The findings obtained from these studies suggest that the relationship between repetition and L2 learning is probably too complex to be sufficiently understood solely in terms of simple causal explanations. It is worth further exploring the benefits of procedural repetition with an orientation towards language pedagogy because classroom-based teaching is usually built up around repeated task performances over several weeks or months. Focus on procedural repetition also encourages researchers to longitudinally explore the effects of repetition on L2 learning, while most of the previous research has only involved limited repetitions (generally two or three) over a relatively short period (generally one day or several weeks).

Table 1. Summary of previous research on oral task repetition

Study	Types of repetitions	No. of repetitions	Length of research	Focus of analysis	Level of analysis
Arevart & Nation (1991)	ER, PR	3	1 day	Linguistic	Group
Plough & Gass (1993)	PR	3	1 week	Interactional	Group
Bygate (1996)	ER	2	3 days	Linguistic	Individual
Gass et al. (1999)	ER, PR	4	1 week	Linguistic	Group
Lynch & Maclean (2000, 2001)	ER	6	1 day	Linguistic, learner perceptions	Individual
Bygate (2001), Bygate & Samuda (2005)	ER, PR	2	10 weeks	Linguistic, discourse	Group, individual
Mackey et al. (2007)	ER, PR	2	1 week	Interactional	Group
Pinter (2007)	PR	3	3 weeks	Linguistic, inter-actional, learner perceptions	Individual
Ahmadian & Tavakoli (2011)	ER	2	1 week	Linguistic	Group
De Jong & Perfetti (2011)	ER, PR	3	4 weeks	Linguistic	Group
Kim (2013)	ER, PR	3	1 week	Language related episodes, learner & teacher perceptions	Group
Fukuta (2015)	ER	2	1 week	Linguistic, attention orientation	Group

Note: ER – exact repetition, PR – procedural repetition

Benefits of writing task repetition: Internal repetition and self-regulation

Before introducing the perspective of complex dynamic systems, we would like to consider the potential benefits of repeating a writing task, which might differ from those of oral task repetition in some significant ways. Manchón (2014a) has drawn our attention to a dichotomy between 'external' and 'internal repetition' introduced by Bygate (2006), to take into consideration the distinctive nature of writing tasks. External repetition is repetition where learners repeat the entire task, which is the design feature adopted in most research on oral task repetition. Manchón (2014a), however, argued that internal repetition, which was originally defined as repetition arising through the processes of a given task "encouraged by the demands of processing the input material and/or of preparing the intended task outcome" (Bygate, 2006, p. 173), may well be more prominent in writing.

Two characteristics of writing, among others, are particularly relevant to the present discussion – the greater availability of production time and the recursive nature of complex cognitive operations. Both characteristics are closely interrelated – the greater availability of time allows writers to engage in deeper cognitive processing such as idea generation, text construction and revision. Thus, learners are likely to be engaged in meaningful and challenging production in a writing task. What is more, these cognitive operations do not occur linearly, but rather recursively (Manchón, 2014b). Thus, in doing a writing task learners often repeatedly engage with each part of the composition process in increasing depth. In other words, internal repetition is essential to the written modality.

Focus on internal repetition illuminates an aspect of individual differences (IDs) in learner agency because varying degrees of internal repetition by each learner are likely to impact on their learning. IDs in learner agency, which have largely been neglected in the oral domain (however, see Lynch & Maclean 2000, 2001 for some aspects of this), would be one of the primary issues in investigating writing task repetition (Manchón, 2014a, 2014b). As one property of learner agency, this study adopts a concept of self-regulation – the degree to which individuals are active participants in their own learning (e.g., Boekaerts, Pintrich & Zeidner, 2000; Heckhause & Dweck, 1998; Schunk & Zimmerman, 2008). Self-regulation is regarded as a key function in writing production. For example, in their well-known models of 'knowledge-telling' and 'knowledge-transforming' Bereiter and Scardamalia (1987) postulated that writers should employ effective use of self-regulatory processes (e.g., analysing problems and setting writing goals) in the more advanced, knowledge-transforming model.

Self-regulation is conceptualised as an open-ended cyclical process of forethought, performance, and self-reflection phases (e.g., Zimmerman, 1998). Particularly important processes in the forethought phase concern goal-setting, which are viewed as guiding principles that learners consciously and intentionally set so as to effectively steer their behaviour in the performance phase (Austin & Vancouver, 1996). In the self-reflection phase, self-regulated learners critically self-evaluate their own learning-related behaviors based on their self-set goals. These attempts should further influence learners during the processes in the cycles that follow in: (a) setting and revising their own goals, (b) monitoring their production, and (c) evaluating their own learning in relation to their own goals. Given such recursive cycles with continuous feedback loops, the developmental patterns of learning are "inherently idiosyncratic to each individual learner" (Nakata, 2016, p. 60).

Understanding task repetition from the perspective of complex dynamic systems

With the cyclical model of self-regulation in mind, let us now consider the longitudinal benefits of writing task repetition from the perspective of complex dynamic systems (CDSs). A CDS is a system that consists of many interacting components (or subsystems) with micro-level interaction between the components that leads to the emergence of macro-level patterns over time (van Geert, 2011). Therefore, a CDS approach assumes that language learning emerges from interaction between various internal (such as linguistic and self-regulatory subsystems) and external influences (such as tasks and a teacher).

Understanding task repetition from the perspective of CDS can offer important insights into L2 learning from a longitudinal perspective. It is not uncommon to observe unexpected and divergent behaviours arising from use of a task in English as a Foreign Language (EFL) classrooms, even when a class consists of homogeneous learners with the same L1 background and equivalent levels of English proficiency. Even if an opportunity for task engagement might not lead to clearly revealing observable differences between learners, each experience of task engagement will leave a trace that could change the latent potential of each learners' language resources. Therefore, when learners are frequently re-engaged in tasks, the same learning conditions are likely to lead to diverse patterns of development over time. This phenomenon can be explained by the property of *iteration* in CDS terms – through repeated operations of the same procedure, the preceding iteration works as input for the next iteration and the result again serves as input for the following iterations. That is, "the present growth level depends on the previous growth level plus the interaction between the resources available at that point" (de Bot, Lowie, & Verspoor, 2007, p. 13). In doing so, repetition of the same operation leads to diverse and often unexpected patterns of change over time.

The emergence of diverse patterns from a task reminds us of the well-known distinction between *task-as-workplan* (i.e., what a teacher or researcher intends or hopes the task would achieve) and *task-as-process* (i.e., what learners actually derive from the task). Breen (1987/2009) argues that this disparity arises because "any learning outcome is the result of a fairly unpredictable interaction between the learner, the task, and the task situation" (p. 334). This is consistent with the CDS view that recognises learning as emerging from the interaction between various internal and external factors. Such divergent effects of task repetition have been reported by Lynch and Maclean (2001): They compared repeated performances by two learners at markedly different levels of English proficiency and found that both reacted to communicative challenges in different ways and benefited in different linguistic dimensions. In other words, task repetition does not necessarily lead to uniform effects but rather to

different benefits for different learners. Such various benefits emerge on the grounds that "any learning outcome will be significantly shaped by *the learners' own perceptions* of all three [learner, task, and task situations]" (emphasis in original; Breen, 1987/2009, p. 334).

Following this perspective that emphasises the contribution of learner agency, we use the concept of *affordance*, which was first proposed by an American psychologist, James Gibson (1979), and has more recently been applied to language learning contexts by van Lier (2000). An affordance is defined as any possible transaction between an individual and the environment, including of course resources in the environment. In the present interest, an affordance concerns the relationship between a learner and a specific context: What becomes an affordance depends on learner agency – what a learner does, what s/he wants and what is useful for her/him (van Lier, 2000, 2004). Thus, what a particular task can afford not only varies across different learners but also within the same learner at different times. Task affordances are therefore unstable and dynamically change through repeated opportunities – learners may follow the task-as-workplan in the first place, but "thereby, make it manageable" (Breen, 1987/2009, p. 334) on the following occasions. In other words, a task sets the initial condition, but because needs and perceptions change from one iteration to another, benefits of task repetition will imply changes in perceived affordances. It can be assumed from this *task-as-affordance* perspective that a task does not directly cause learning. Rather, interaction between a learner and the task constructs and constrains the learning affordances (Larsen-Freeman & Cameron, 2008).

The present study

The present study is aimed at exploring this dimension of task-as-affordance by researching changes in L2 writing and self-regulatory processes over time through writing task repetition. Nitta and Baba (2014) investigated what effects the repetition of a timed-writing task (in terms of both exact and procedural repetitions) would have on the various linguistic measures of two different EFL classes over one year to explore the benefits of writing task repetition over an extended period. Group-level analysis indicated that the effects of exact repetition were very limited, whereas procedural repetition had marked effects on lexical and syntactic measures in both classes. To further understand the effects of writing task repetition, Baba and Nitta (2014) explored patterns in L2 writing development by observing two focal students, who experienced discontinuous changes called 'phase transitions' in their writing fluency through 30 repetitions. The findings obtained from these longitudinal studies also revealed that the benefits of writing task repetition differed from those of oral tasks. While exact repetition led to limited effects, procedural repetition of this writing task

for an extended period encouraged individual learners to develop various linguistic dimensions and to follow their own unique developmental paths. While acknowledging that repetition made significant contributions to the emergence of distinctive ways of L2 writing development, Baba and Nitta (2014) noted, from their post-writing reflections, that the two focal students expressed individual enjoyment and active engagement in the task. In other words, learners' agentic attitudes towards the task might also play an essential role in writing development.

The present study is based on our previous investigations and is aimed at searching for the internal dimensions of writing task repetition by focusing attention on learners' interactions with the task in terms of IDs in learner agency, in particular their efforts at and conscious operations of self-regulation. To understand iterative functioning of self-regulation, we analysed learners' self-reflection in L1, which was carried out after each engagement in a timed-writing task. We also examined qualitative changes in their writing in terms of the patterns and complexity of language produced by students, and quantitative changes, in interaction with their self-regulatory cycles. The following research question guided us in our investigations: How did self-regulatory processes influence changes in L2 writing through learners' repeating specific writing topics (i.e., exact repetition) and the same procedure (i.e., procedural repetition) over thirty weeks?

Method

Background

We collected our data from an EFL classroom composed of 26 first-year English majors at a Japanese university. The class was a compulsory course offered for one academic year (30 weeks) with the objective of integrating four English skills through engagement in various project-based activities (There was a two-month summer break between Weeks 15 and 16). The task presented in the following section was provided weekly at the end of each class. As the course was taught by one of the authors, we were able to gain significant insights into individual students' learning behaviours on a regular basis (see below *Analysis* for advantages of this method). All the students had little experience of writing in English before entering university, and even in university, they had little chance to write in English outside the classroom. The class was a group of similarly proficient students based on the results of an English test called the Test of English for International Communication (TOEIC). The class average was 414.43, which is considered to be around the A2 Threshold of the Common European Framework of Reference for Languages (Council of Europe, 2001).

Task and procedure

The study involved two conditions: exact repetition (i.e., repetition of the same topic and procedure) and procedural repetition (i.e., repetition of the same procedure but different topics). Students wrote a composition in English on a topic of their choice for 10 minutes. The writing task was meaning-oriented, and learners were directed to write their own thoughts about the topic by focusing on fluency and content in their writing as their first priority.

The topics were selected and adapted from a list of essay topics for the Regents' Test (Weigle, 2006). They were given a list of three topics each time (e.g., *'Name your favourite game or sport and explain why you find it enjoyable?', 'What foreign country would you like to visit, and why would you like to go there?'*, and *'What was your most favourite subject in high school? Explain why'*). The aim of offering three alternatives was to compensate for differences in the students' individual experiences and preferences. As suggested by the example topics above, the writing tasks shared some level of procedural similarity while focussing on different topics; this was the case throughout the thirty weeks of the academic year. The same list of three topics was used two weeks in a row. Students chose a topic, and were required to write about the same topic the next week without looking at their previous compositions (i.e., exact repetition). They then received another new topic list the next week (i.e., procedural repetition). Much like in the real world, they were allowed to use a dictionary. It should be acknowledged that even though learners were required to write about the same topic under the exact repetition condition, they were not instructed either to rewrite the same content or to write about different content concerning the same topic – thus, task repetition for this study can be regarded as 'recycling' or 'retrying' (Lynch & Maclean, 2000, 2001).

After they had completed the writing part, students counted the number of words they had written, recorded this, and then filled out a reflection sheet in their L1. There were three sections on the reflection sheet: (1) comments on that day's composition in terms of (a) grammar and vocabulary, (b) organisation and expression, (c) content, and (d) writing processes and strategies, (2) comments comparing that day's composition with previous compositions, and (3) goals for future writing. It usually took about 10 minutes for the students to complete the reflection sheet. These self-reflective opportunities might have influenced students' learning, as has been suggested by psychological research on reflection, which states that learners could turn experience into learning through self-reflective practice (e.g., Boud, Keogh, & Walker, 1985). More specifically, research on self-regulation demonstrates that opportunities for recording word counts and reflecting on their writing could encourage their self-regulatory processes (Zimmerman & Risemberg 1997).

Completed compositions and self-reflections were checked by the teacher/researcher and returned with some feedback comments the following week. The feedback was only aimed at encouraging positive attitudes towards writing. Linguistic correction was not offered to keep the students' attention primarily focussed on meaningful aspects of their writing.

Analysis

We adopted a case-based method taken from the perspective of CDS (Byrne & Regin, 2009), as we had done in our previous studies (Baba & Nitta, 2014; Nitta & Baba, 2015). Traditional research has centred on "the notion of the variable as the causal agent in the world", while the present methodology focused on cases as complex systems, in which "trajectories and transformations depend on all of the whole, the parts, the interactions among parts and the whole, and the interactions of any system with other complex systems among which it is nested and with which it intersects" (Byrne, 2009, p. 2). While various system levels can be set as cases (e.g., a class and a school), this study regarded individuals as cases who changed as a consequence of interaction between external (e.g., a task) and internal subsystems (i.e., self-regulatory processes).

Because it was not practically possible to examine all interacting factors, the CDS approach attempted to explain phenomena *retrospectively* by focusing on limited numbers of components and their patterns of change, which were likely to operate in significant ways and emerge within other contexts. In other words, "by tracing back the reasons why the system has ended up with a particular outcome option we produce a retrospective qualitative model of its evolution" (Dörnyei, 2014, p. 85). Our teacher-researcher position seemed to have a great advantage in this respect because our regular observations of students' behaviours and learning made us aware when they learned (and when they did not learn) and what factors might have influenced their learning.

Based on our retrospection, we decided to focus on self-regulatory functions presented in the students' post-writing reflections because we noticed that the quality and quantity of their self-reflections seemed to be related to their L2 writing development in significant ways. Some students were engaged in writing their self-reflections elaborately, while others seemed to just fill in the post-writing reflection sheet. These differences might only have had marginal impact on changes in their L2 writing in the short term, but such different degrees of task engagement were likely to influence the quality of their writing over time. For example, Storch (2008) found, by investigating the metatalk of learners in pairs, that learner engagement in a task ranged from elaborate to limited and that elaborate engagement was more facilitative of learning. Such different qualities of learning in the self-regulation literature are conceptualised as 'skilful' and 'naïve' – the former learners benefit from positive cycles of clear goal

setting, careful monitoring and critical self-evaluation of their learning, whereas the latter are trapped in vicious cycles of unspecific goal setting, limited extent of monitoring and avoidance of self-evaluation (Zimmerman, 1998, 2002).

Working backwards and pinpointing "the principal factors that have led to specific settled states" (Dörnyei, 2014, p. 85) of skilful and naïve self-regulated learners, we focused on two students, Yuto and Taka (both were males and their names are pseudonyms), who had distinctive differences in terms of English writing and self-reflective comments. Yuto was selected as a typical case who demonstrated an average amount of reflection and similar patterns of writing changes, while Taka was selected as an unsuccessful case who demonstrated very limited self-reflection and writing development (see findings below for details; these are also presented elsewhere – see Nitta & Baba, 2015 for other cases). The multiple-case method aims to illuminate possible factors leading to very different outcomes and contribute to our understanding of task repetition research. In this respect, it was suggested by van Lier (2005, p. 198) that "if two cases provide apparently contradictory information about a certain issue … this contrast can provide much food for thought and future research, thus being of great benefit to the field".

To analyse both students' reflections written in L1, we first looked at the number of Japanese characters as a quantitative index of the degree of the students' task engagement, or 'a state of heightened attention and involvement' (Philp & Duchesne, 2016, p. 51) in the present writing task. Because this section was written in their L1, language problems did not hinder their writing, and the amount they wrote was likely to indicate the extent to which they engaged with the tasks. Because this measure was obtained each time, it also reflected changes in their degree of engagement over the year.

Because such a gross measure only captures some aspects of behavioural engagement, we also identified their self-regulatory processes represented in their self-reflection. We coded the reflections according to three key self-regulatory processes to analyse the qualitative aspects of L1 reflections: goal-setting, self-observation and self-evaluation. Of the three sections on the reflection sheet (see *Task and procedures* above), 'goal-setting' was coded from the third section (i.e., 'goals for future writing'), whereas 'self-observation' and 'self-evaluation' were coded from the first two sections (i.e., 'comments on that day's composition' and 'comments comparing that day's composition with previous compositions'). Self-observation in the present analysis concerned students' impressionistic descriptions of their thoughts or actions, which often contained somewhat superficial descriptions of their own writing. Self-evaluation, on the other hand, required evaluative comments that involved several cognitive processes such as specifying, reasoning, analysing, or comparing reflections.

To analyse their L2 writing we examined fluency and syntactic and lexical complexity using Coh-Metrix (Graesser, et al., 2004, available at <http://cohmetrix.

memphis.edu/cohmetrixpr/index.html>). Writing fluency is concerned with how fast one writes for ten minutes – we used text length (TL, i.e., number of words per composition) as an indirect measure of speed. We used average sentence length (ASL) as a syntactic complexity measure, which was in line with Ortega (2003), whose research synthesis on syntactic complexity in L2 writing had found that it differed according to L2 proficiency levels. We used the Measure of Textual Lexical Diversity (MTLD) for lexical complexity, because this is not affected by text length (McCarthy & Jarvis, 2010). Accuracy measures were excluded from the present analysis because focus on accuracy might have kept students from producing as much meaningful writing as possible. Exclusion of accuracy was also a practical necessity with a database of more than 700 compositions from 26 students, given that Coh-Metrix did not track accuracy.

Results

Yuto

Yuto was regarded as a representative case in terms of the production of reflective comments. His total reflection was composed of 3,324 Japanese characters, which was roughly equivalent to the class average of 3,330. The change in the degree of reflection each week is provided in Figure 1 depicting the altitude lines (van Geert & van Dijk, 2002) of the whole class, which plot the 75th, 50th, and 25th percentiles. The graph allows altitude and steepness to be visually inspected while its primary focus is on changes in individuals, thus enabling individual changes to be investigated in comparison to the group-level transitions of states and the identification of variation ranges. For example, a shift from around the 25th percentile line to above the 75th percentile line can be interpreted as evidence for development over time relative to other students. Despite occasional drops, the values for Yuto continued to be plotted around the 50th line of the class values for most of the time.

Figure 2 plots changes in the employment of each self-regulatory process throughout the year. This allows us to explore Yuto's engagement in the task in more depth. The figure indicates that he tended to employ self-evaluation (i.e., accounts involving specifying, reasoning, analysing, or comparing) more frequently (68 times) than self-observation (i.e., superficial and impressionistic accounts of his thoughts or actions (41 times)), while he continued to set his own writing goals throughout the year (36 times).

The following provides several examples of Yuto's specific accounts, which were coded as self-evaluation:

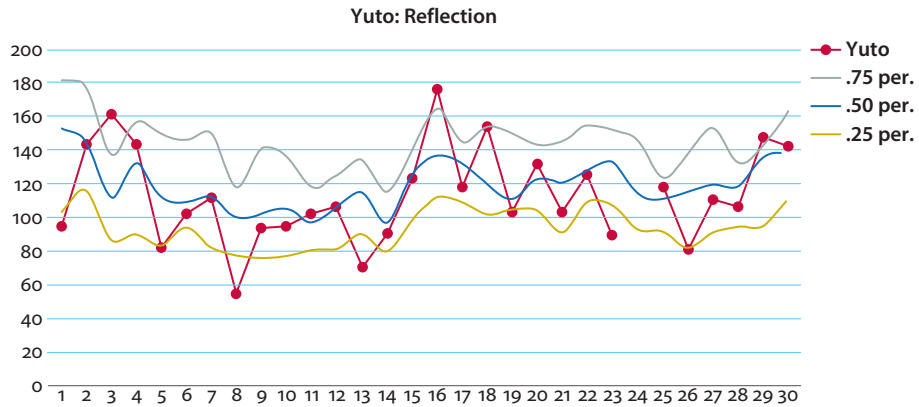

Figure 1. Changes in degrees of self-reflection for Yuto with altitude lines of class

- *'I was repeatedly using an "I-pattern", so I should use other ways like an "It is something that" structure.'* (Self-evaluation: Week 1)
- *'It was easy for me to predetermine certain forms like "I want to…", "because" and "for example".'* (Self-evaluation: Week 16)
- *'I made some time to intentionally check the language, and I found a lot of mistakes on copular verbs and use of capital letters.'* (Self-evaluation: Week 29)

These examples, which are identified at different points in time, suggest that Yuto tended to self-reflect on his writing in a specific and critical manner throughout the year.

Figure 3 plots changes in the linguistic measures of his L2 writing (i.e., TL, ASL, and MTLD) with the altitude line graphs over the thirty weeks. The changes in his writing indicate an incremental trend in TL in the first half and relative stagnation in the second half of the year. This trajectory suggests his focus on improving writing fluency from the beginning to the mid-point of the academic year. The latter half of the year, on the other hand, seems to suggest that he had instead focused on improving the other linguistic measures; ASL and MTLD demonstrated relative development throughout the year, whereas these values tended to be more prominent during the second half.

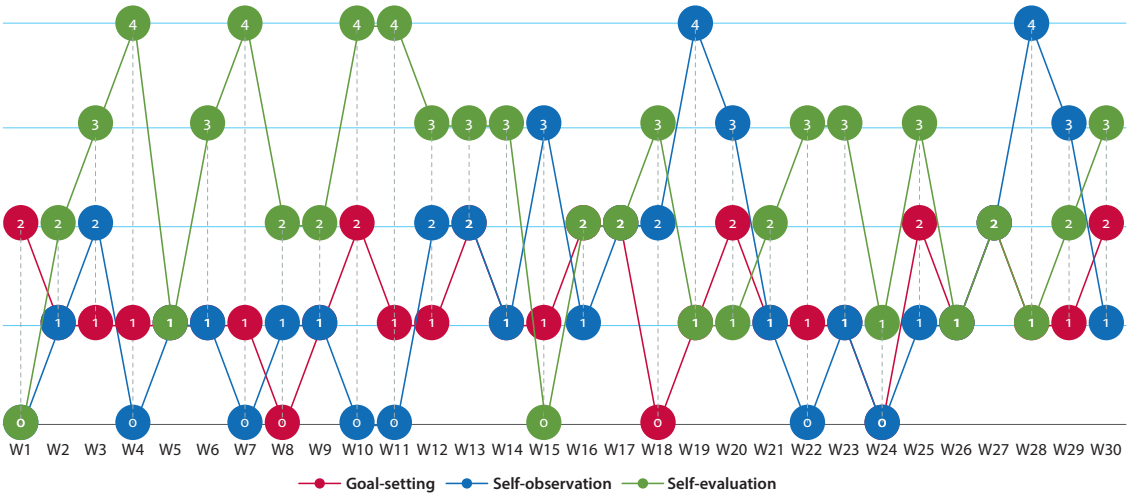

Figure 2. Changes in self-regulatory process: Yuto

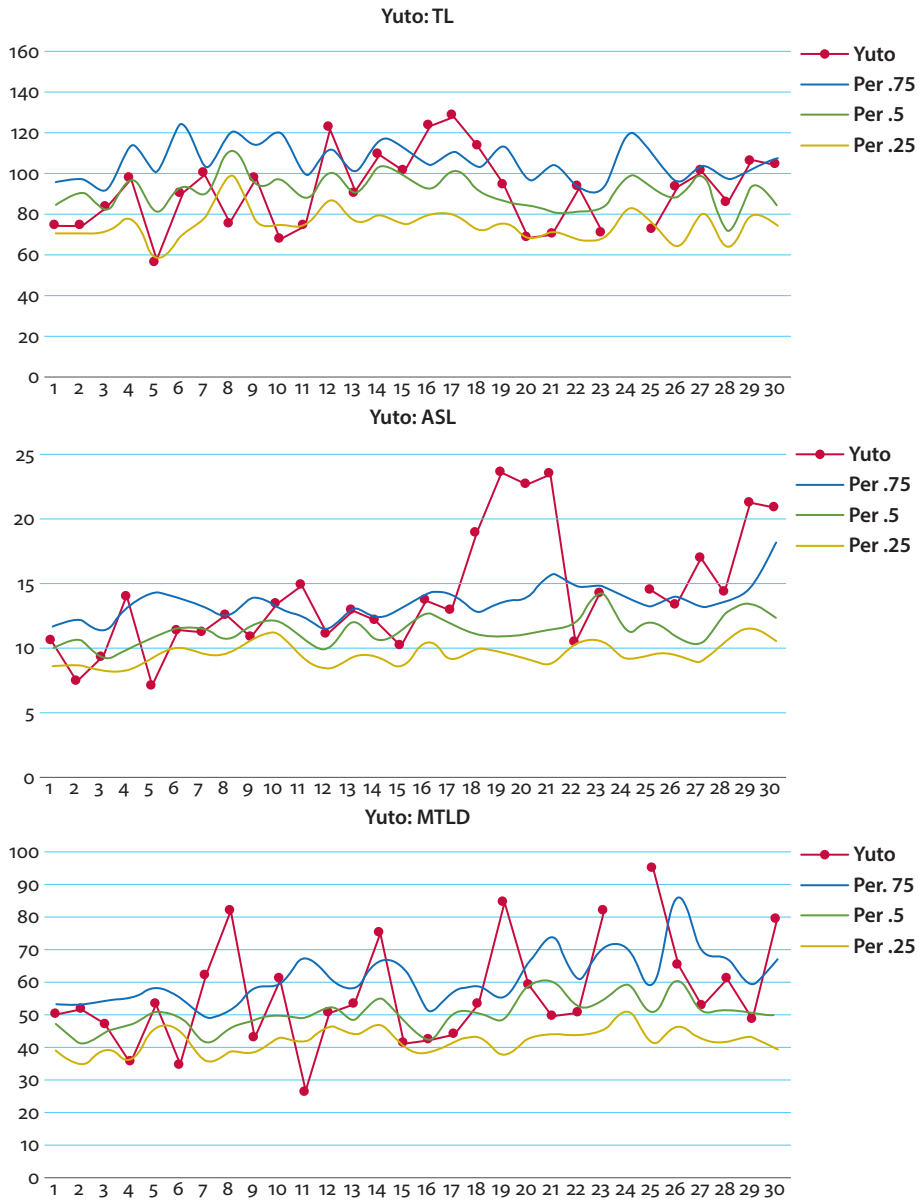

Figure 3. Changes in L2 writing measures: Yuto
Note: TL – Text Length, ASL – Average Sentence Length, MTLD – Measure of Textual and Lexical Diversity

Below, we will provide a few examples of Yuto's writing – from around the beginning and the end of the year – to illustrate the changes we observed. The Week 2

composition consisted mostly of simple sentence structures involving 'I' as a subject, which were loosely connected by using coordinating conjunctions (e.g., 'but' and 'so'):

> I want to go to Switzerland. Because it has glamorous clear air, beautiful mountains. Other reason, I started to study German in university, so I want to use it to communicate with local people who speak German. But I'm not good at German of well. So I have to study about German history, language, and culture. Other reason, it is place of Anime "Heidi". So I want to go actual place of Heidi.
>
> (Week 2: Repeated; TL – 74, ASL – 7.4, MTLD – 51.5)

This composition provides three reasons, the second of which (i.e., 'to study German') is relatively elaborate. However, because he simply used the same conjunction ('so') three times to link reasons to intentions, the whole composition created an impression of just listing ideas without sufficiently elaborating on the topic.

In contrast, his final composition includes more elaborate ideas and structures. Other nouns and pronouns were also used as subjects, and many of the sentences were developed to provide detailed information by using a wider range of connectors:

> Suppose I was asked to teach a subject, sport, or activity of my choice, I would teach football to junior high students, because I have played football for about 10 years, and I was taught a lot of things by my coach when I was junior high student. When the player is junior high student, think themselves about their play, team, …etc. It is different from elementary students. Then, they need someone's advice, so if I could help them, I would like to teach my skill or advice my idea. Thus I wish they had get some beneficial things to enjoy playing football.
>
> (Week 30: First time; TL – 104, ASL – 20.8, MTLD – 79)

In comparison with the week 2 example above, this composition employs a wide range of connectors (e.g., 'suppose', 'because', and 'when') and provides detailed reasons for his choice – teaching football to junior high school students – by referring to his own experience.

A natural question that arises is what made his writing improve in such ways between the two points in time. A noticeable characteristic observed in his writing between the beginning and the end of the year was his effective use of opportunities to repeat the same topic. Table 2 lists examples of the first and repeated writing with his reflective comments in Weeks 3 and 4.

In the exact repetition condition (Week 4) he upgraded his writing in terms of both quantity and quality, suggesting that for him this condition offered meaningful opportunities. In addition to slight quantitative increases in the measures of TL and ASL, closer examination suggests some clear evidence of qualitative improvements. The repeated composition (Week 4) included significantly greater amounts of information and organisation – that is, in the second version he moved 'music' into the

first paragraph, which dealt with solitary activities; in contrast, he provided a general comment about the value of staying with friends in the second paragraph.

His self-reflection data reveals how he tried to improve his writing through repetition. The Week 3 composition (in the left column) includes three ideas in each

Table 2. Examples of writing and self-reflection: Weeks 3 and 4 for Yuto

L2 Writing ('When you are unhappy, what do you do that usually makes you feel better?')	Self-reflection (translated from L1)
Week 3	
When I am unhappy, I usually make me feel better to do a variety of things. *One of them is sleeping for a long time. When I sleep, I think when I next wake up, I feel good.* *Other of them, I enjoy playing sports with my good friends. After enjoy playing it my feeling is fresh. And I like sports very much. So it is so good thing.* *Also I listen to cheerful music. it will make me feel a lot comfortable.* (TL – 83, ASL – 9.22, MTLD – 35.5)	<u>Grammar & vocabulary</u> 1. I tried to use 'It makes me'. (S/E) 2. I couldn't recall adjectives and adverbs immediately without consulting a dictionary. (S/E) 3. While writing, I noticed sentences were incomprehensible. (S/E) <u>Organisation & expression</u> 4. I thought I could increase the number of words if I would write more specific examples and my experiences. (S/E) <u>Writing processes & strategies</u> 5. The topic wasn't appropriate for me, so I found it difficult to write this time. (S/E) <u>Goals</u> 6. I want to plan a little more before starting to write. (G/S)
Week 4	
When I was sad, I sometimes watch sad movie or listen to sad music, after then, I watch comedy movie or merry music. After do this, I refresh my feeling, and forget sad things. So it is my solution when I am unhappy. *Another thing my solution, it is enjoy playing my friends, for example, go to karaoke, playing football, and chatting. It makes me fine, when I am unhappy, I think. It is not good that I am alone, so I am helped by my good friends. Therefore, my good friends are very important for me.* (TL – 97, ASL – 13.9, MTLD – 35.5)	<u>Grammar & vocabulary</u> 7. Because I didn't use a dictionary, I could manage to describe my own ways instead of using difficult vocabulary. (S/E) <u>Organisation & expression</u> 8. Using 'First', 'Second' styles seems to make my writing more readable. (S/E) <u>Comparison with previous compositions</u> 9. Because I wrote in a specific manner, I could increase the number of words. (S/E) 10. Not only because I planned what to write before starting to write but because this is the second time of writing the same topic, I could write more smoothly. (S/E) <u>Goals</u> 11. I think I'll try to use simpler vocabulary that I understand more frequently. (G/S)

Note: TL – Text Length, ASL – Average Sentence Length, MTLD – Measure of Textual and Lexical Diversity; G/S – goal-setting, S/E – self-evaluation

paragraph where adjacent sentences are only loosely connected without the ideas being sufficiently developed or the language being carefully monitored. His awareness of flaws in his composition is revealed in his self-reflection (in the right column). For instance, he noted that his sentences were rather "incomprehensible" (No. 3 in Table 2); thus, to overcome this problem, he set a goal of planning before starting to write (No. 4). He also understood that writing about specific examples and his own experiences would contribute to increasing the total number of words (No. 4). He evaluated himself in the following week (Week 4) and stated that he had successfully achieved his self-set goals by planning before writing and adding specific events (No. 10), and that these attempts contributed to improving his quality of writing. As reflected in his account (Nos. 8 and 9), the Week 4 composition demonstrated some improvements in terms of sentence connections and idea elaboration.

Table 3 presents another example of improvement in his writing, which came as a result of repeating the same topic. Like in the Week 3 composition, he seemed to have focused on translating his ideas into English without thinking carefully about the language when he was engaged in the Week 13 composition.

In addition to the increases in TL and MTLD, the second version (Week 14) included improved sentence and paragraph organisation. Many of the connectors were not effectively connected to individual sentences in the Week 13 composition, while the whole composition was organised more coherently by employing comprehensible sentences and coherent connections in the Week 14 composition. Furthermore, the second version was improved in terms of its content. The argument is rather vague in Week 13, while the Week 14 composition has clearer principles, for example by his stating: "*However, it is very important to cooperate with other people and aim at the same objectives*".

Yuto observed in his self-reflection (in the right-hand column) that he was unable to monitor his grammatical use in the Week 13 composition because he placed too much emphasis on thinking about the content (No. 1 in Table 3). He self-reflected in week 14 that he was again not able to focus on grammar sufficiently due to his attempt at making longer sentences (No. 7), although he expressed satisfaction about this writing process compared to his writing in the previous weeks (No. 9).

To summarise, because the task instructions emphasised writing fluency through the counting of written words, Yuto seemed to have focused on increasing the text length in the beginning weeks whereas he increasingly attempted to focus attention on formal aspects during the final weeks. From the very early weeks, he reflected on concerns about his writing (e.g., "I only used the 'I-pattern' in most sentences" [Week 2] and "My grammar was clumsy!" [Week 7]). Then, he became aware of grammar while writing as reflected in "I was a little conscious about my grammar" (Week 8) and "I was very conscious about grammar this time" (Week 12). In addition, while he focused on the sentence-level structure in the beginning weeks, he became more aware

of the paragraph-level structure of his writing. He reflected on his sentence connections more explicitly and critically self-evaluated his attempts (e.g., "I only used the same way connecting sentences" [Weeks 28 & 29], and "Because I tend to always use 'so', I want to be able to use other connectors" [Week 30]). In other words, both opportunities of exact and procedural repetition contributed to improving his writing.

Table 3. Examples of writing and self-reflection: Weeks 13 and 14 for Yuto

L2 Writing ('What advice would you give to a student just beginning high school?')	Self-reflection (translated from L1)
Week 13	
A student just beginning high school is advised to cooperate with classmate, teammate of club activity, and so on. To cooperate and aim to same direction is very good thing. In this way, it has a lot of conflicts or other problems, and sometimes happens quarrel. However, when it is cleared to good way. You can get a lot of important things of your life. So cooperate with people is very good things. (TL – 90, ASL – 12.9, MTLD – 53.1)	Grammar & vocabulary 1. Because I spent too much effort on thinking about the content, I couldn't pay attention to grammar while writing. But I tried to make each sentence longer. (S/E) 2. I want to increase the number of connectors I can use. (S/O) 3. I tried to write passive sentences. (S/E) Content 4. Because the topic was too difficult, I couldn't come up with any ideas. (S/E) Comparison with previous compositions 5. I think it was easier to write than the last topic. (S/O) Goals 6. I want to write a greater amount of words. (G/S)
Week 14	
I want to some advices for a student just beginning high school. It is about cooperating with a lot of people, for example, classmate, teammate in club activity, or any other same person. Because after graduated high school, we have seldom chance to cooperate with other people. However, it is very important to cooperate with other people and aim to the same objective. So when a person belongs in high school, the person should takes part in events like it needs cooperation of a lot of people, for example, School Festival, tournament of club activity. (TL – 109, ASL – 12.1, MTLD – 75)	Grammar & vocabulary 7. Because I focused on making each sentence longer, I couldn't write while paying attention to grammar. (S/E) Organisation & expression 8. It was easier to write by making the whole into several parts. (S/E) Writing processes & strategies 9. This time I came up with many ideas I wanted to write, so I could write smoothly. (S/E) Comparison with previous compositions 10. It became easier to write rather than before. (S/O) Goals 11. I want to be able to write more sophisticated sentences. (G/S)

Note: TL – Text Length, ASL – Average Sentence Length, MTLD – Measure of Textual and Lexical Diversity; G/S – goal-setting, S/O – self-observation, S/E – self-evaluation

Taka

Another student, Taka, is presented here as a rather unsuccessful case who demonstrated very limited development in writing. A lack of improvement in his writing through task repetition might be regarded as a consequence of his limited engagement. His total amount of reflection was 1,927 Japanese characters, which was far lower than the class average of 3,330. The changing values of his reflection each week are provided in Figure 4, in which Taka's reflective comments are always located on or below the 25th values of the class throughout the year. This suggests that his lack of motivation towards the task was consistent.

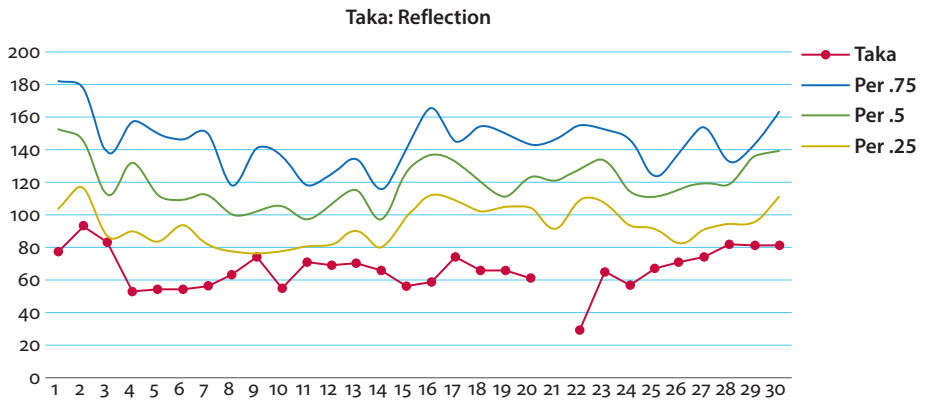

Figure 4. Changes in degrees of self-reflection for taka with altitude lines of class

Figure 5 suggests that Taka's self-regulatory processes largely involved self-observation (133 times) with very limited use of self-evaluation (two times), while goal-setting was constantly identified (26 times) throughout the year.

Most of his self-reflective accounts lacked elaborate descriptions and detailed explanations. This is illustrated in the following excerpts:

- *'I was writing while recalling past experience.'* (Self-observation: Week 2)
- *'I wrote about one person.'* (Self-observation: Week 14)
- *'I could explain well.'* (Self-observation: Week 28).

Figure 6 shows changes in the three measures of L2 writing, which indicates no clear developmental trends. Despite occasionally high values (i.e., Weeks 4, 13, 15, and 30), the measure of TL does not suggest a clear developmental pattern. In a similar vein, although the ASL values are often plotted on and above the 75th percentile line throughout the year, they do not indicate any incremental trends, and the MTLD values tend to be plotted between the 25th and 50th percentile lines throughout the year.

The compositions he produced at the beginning and the end of the year give an impression of the stability of his writing quality during the observation period.

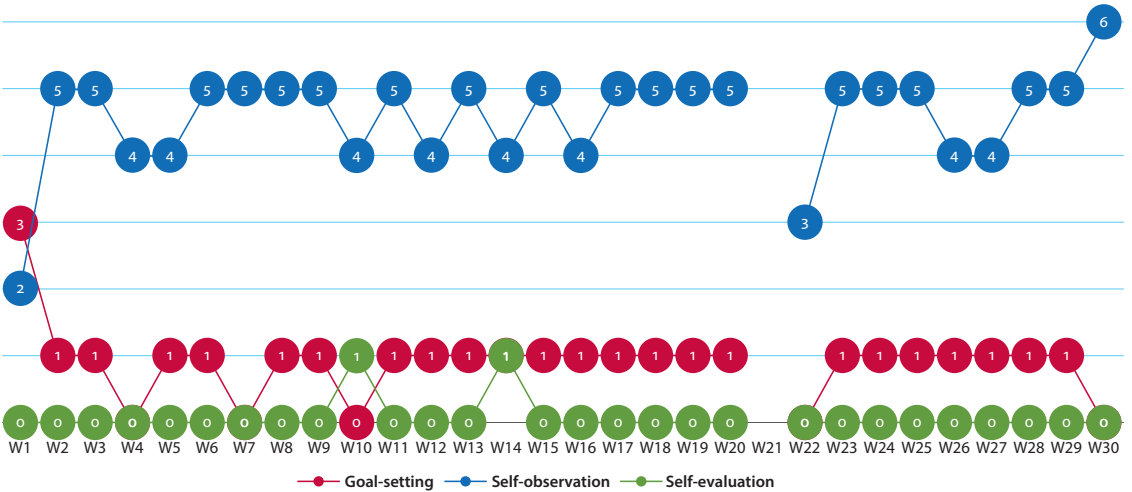

Figure 5. Changes in self-regulatory process: Taka

His Week 2 composition, in which sentences are rather simply connected, might be deemed slightly better than Yuto's beginning composition (presented above) in terms

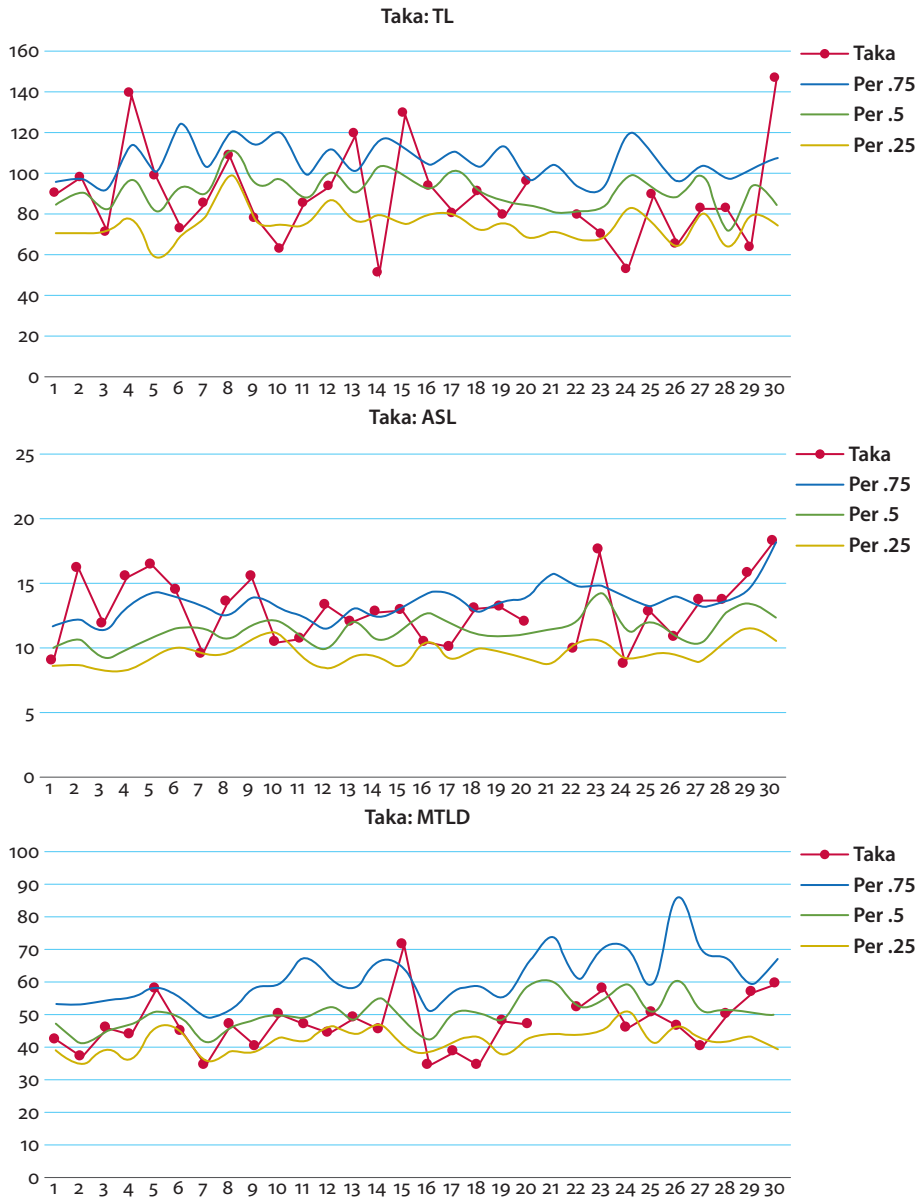

Figure 6. Changes in L2 writing measures: Taka
Note: TL – Text Length, ASL – Average Sentence Length, MTLD – Measure of Textual and Lexical Diversity

of his elaboration of ideas; this composition seems to reflect his attempt to develop ideas from sentence to sentence, rather than just listing reasons.

> At first I want to introduce my favorite sport. My favorite sport is basketball and soccer. The reason why I like those two sports is I have been playing for long time of those two sports. Second reason is I like to watch those sports on TV. Many cool players are playing. And I thought I wanted to play like them! And when I kept practicing and practicing I found it was so enjoyable. But now in university I am not playing sport. So sometimes I will play sport in private. I think everybody will enjoy sport.
>
> (Week 2: Repeated; TL – 97, ASL – 12, MTLD – 37)

Despite Taka's better quality of writing than that of his classmates, significant progress could not be observed in the following weeks. The compositions below were produced in the latter half of the year, which do not show any qualitative changes, as shown by the stagnant trends in the linguistic measures (Figure 6).

> I want to drive a car that looks like a sports car. I will choose without looking. Of course function is important but I think out looking is more important. From this, I am now asking my father to buy me a nice out looking car. But I don't want to cost a lot of money for the car, so I am finding with the magazine and with the Internet.
>
> (Week 19: Repeated; TL – 79, ASL – 13, MTLD – 48)

> If I could have a conversation with a famous person, I would like to meet Michael Jordan. He was a basketball player and he plays basketball very good. When I was 12 years old I saw Michael on TV and he made me playing basketball at junior high school. At the school I was first time playing basketball but I study hard basketball.
>
> (Week 29: Repeated; TL – 63, ASL – 16, MTLD – 57)

The lack of change in his writing can be attributed to the poor quality of his self-reflections written in L1. Not only did he produce fewer reflective comments than the class average (Figure 4), the qualitative analysis also suggests that most of his comments were largely general and superficial. Taka repeatedly set goals in his reflections (see Figure 5) but they were often fragmented and not clearly connected to his reflections in the following weeks. For example, he set form-focused goals in the beginning weeks (e.g., '*I want to remember grammar*' [Week 1], *'I want to write sentence structures accurately.*' [Week 2]), but his following observations were not clearly linked to these goals but were rather concerned with writing fluency (e.g., *'I tried to write fluently without stopping*' [Week 2], *'I tried to write smoothly*' [Week 3]). Then, his goals switched to a focus on fluency in the following weeks (e.g., *'I want to increase the number of word*' [Week 3]), which might be a consequence of his previous observations. Soon, however, his goals started changing again, for example, *'I want to be careful of sentence structures next time*' (Week 5), *'(My next goal is) To remember necessary vocabulary*' (Week 6), and *'I want to write a lot for the next topic'* (Week 8). On the whole,

his goal-setting seemed to be impulsive, and loosely connected to his actual writing. His lack of critical and elaborate self-evaluation created the impression that he merely wanted to get the task (both repetitive writing and self-reflection) done.

As a result, he did not seem to make any clear attempts to exploit the task repetition to improve his writing. Table 4 shows examples of his first and repeated compositions (the left side column) and his self-reflections (the right side column) in Weeks 13 and 14.

Table 4. Examples of writing and self-reflection: Weeks 13–14 for Taka

L2 Writing ('What type of movies do you prefer? Why do you like them?')	Self-reflection (translated from L1)
W13	
I like watching movies. In the movie, I like action movies. Because action makes me exciting and it's so much interesting. I have DVDs almost 300 discs, from this you will know how much I like movies. I don't like watching movie in a movie theater I always watch movie at my house with my DVD player. I think movies are liked from everybody. Today I am going to buy DVD, and I am going to collect DVD more and more. And also I like lent my DVD to my friend. I am hoping that my friend will be happy with that. One of the movie called "Mist" what I saw at my house was so interesting. (TL – 119, ASL – 12, MTLD – 49)	<u>Grammar & vocabulary</u> 1. I wrote what I thought. (S/O) [Organisation & expression] 2. I wrote what I thought. (S/O) <u>Content</u> 3. I wrote about myself. (S/O) <u>Writing process & strategy</u> 4. I tried to write smoothly. (S/O) <u>Comparison with previous compositions</u> 5. I think it was easier to write than before. (S/O) <u>Goals</u> 6. I want be able to write more smoothly. (G/S)
W14	
My most favorite movie is Action movie. I like, because his acts are very good and I have a lot of DVD of him. He is very kind person but when he fights in the movie he is very strong. Everyone likes him and he is a super star in China. (TL – 51, ASL – 13, MTLD – 45)	<u>Grammar & vocabulary</u> 7. The grammar was difficult, so it may be inaccurate. (S/E) <u>Organisation & expression</u> 8. I could write what I thought. (S/O) <u>Content</u> 9. I wrote about one person. (S/O) <u>Writing process & strategy</u> 10. I wrote what I thought. (S/O) <u>Comparison with previous compositions</u> 11. I couldn't write as well as before. (S/O) <u>Goals</u> 12. I'll increase the number of words next time. (G/S)

Note: TL – Text Length, ASL – Average Sentence Length, MTLD – Measure of Textual and Lexical Diversity; G/S – goal-setting, S/O – self-observation, S/E – self-evaluation

In terms of the three linguistic measures, Taka's TLs noticeably decreased from Weeks 13 to 14, and his ASL and MTLD indicated no incremental trends. In line with these quantitative results, the ideas presented in the repeated composition (Week 14) became much more simplified and lacked elaboration about the topic – action movies – in comparison to those presented in the first composition (Week 13). That is, although he developed the idea of action movies to a certain extent in the first composition, for example, by comparing movies he watched at a theatre and at home, he only provided some indication of actors without identifying them by name in the second version.

Reduced amounts of TL might be indicative of his reluctance towards having to retry to write on the same topic again. His lack of motivation to engage in exact repetition was also suggested by the quality of his self-reflection presented in the columns on the right. We identified 'self-observations' (indicated as SO) in most of his accounts in Weeks 13 and 14. In contrast to Yuto's specific accounts, Takas' are largely impressionistic and superficial. For instance, he repeatedly stated "I wrote what I thought" (Nos. 1 and 2 in Week 13, and Nos. 8 and 10 in Week 14) without indicating what he actually thought. In other words, he appears to have thought superficially and to have given up digging deeper into his thinking processes. This suggests that he was reluctant to write about the same topic again and did little more than fill in the task sheet without carefully monitoring and reflecting on his own writing.

Space limitations do not allow further examples, but these aspects of his accounts were typical of his self-reflections throughout the year, as indicated by changes in his self-regulatory processes (Table 5 above). Thus overall, the exact repetition conditions did not encourage Taka to develop his writing skills, in clear contrast to Yuto's attempts to improve L2 writing under the same conditions.

Discussion

The changes (or lack of changes) in the quantity and quality of the two students' written output and self-reflections suggest that their motivation toward the writing tasks played an important role in the performance and development of their writing skills. When a task is repeated over an extended period, self-regulatory processes seem to influence learners' L2 writing development to a decisive extent. Research on self-regulation (e.g., Zimmerman, 2002) suggests that skilful self-regulators (like Yuto) seek out opportunities to self-evaluate their learning efforts, whereas naïve self-regulators (like Taka) are oblivious to them or actively avoid these opportunities. The qualitative analysis of the two students' self-reflections points out the importance of self-regulatory cycles. That is, their self-evaluations of goal achievements led them to re-evaluate the feasibility and desirability of their goals, which inevitably influenced their task

engagement in subsequent writing opportunities. These cycles of self-regulation indicate the emergence of learning as a consequence of repeated engagement of writing and self-reflection. The result of the preceding iteration is taken as input for the next iteration and produces certain developmental patterns in L2 writing.

The present findings are suggestive of different affordances of task repetition in the written modality depending on learner agencies. Our previous study indicated few effects of repeating the same writing topic (i.e., exact repetition) in terms of quantitative changes (Nitta & Baba, 2014). Despite limited changes in the quantitative measures, the present qualitative analysis demonstrated the values of writing task repetition in the case of Yuto, but it should be noted that these benefits were not found in the oral-based research. As indicated in our analysis, he tried to write about the same topic with different goals in mind. In other words, every time the topic was recycled, the task affordances changed. What is more, task affordances not only changed within the same individual at different points in time but between different individuals. That is, the exact repetition, which was regarded as providing valuable learning opportunities to Yuto, was superfluous for Taka. The latter was reluctant to write about the same topic repeatedly, as was suggested by the drastic decrease in the length of his texts.

The interaction between learner agency and L2 writing development made aware of the nature of 'complete interconnectedness' in which "all parts are connected to all other parts" (de Bot & Larsen-Freeman, 2011, p. 10) from the perspective of CDS. That is, a learner and a task are closely connected and reciprocally interact as part of a complex dynamic system. The design features of task-as-workplan are supposed to be fixed – a task is externally located and learners have no control over it. However, our analysis of the two cases suggests that Yuto and Taka were in a position to control the task – in terms of how they interpreted the task and were engaged in it (cf. Coughlan & Duff, 1994). For example, they could select the topic they would write about and elaborate on the content that they thought would be important and meaningful. In other words, language learning emerged as a result of learners' interacting with the task and their agentic attitudes towards the task.

These interpretations allowed us to reconsider how learning takes place – or did not take place as in the case of Taka – through task repetition. Following the work of van Lier (2000, 2004), understanding tasks-as-affordance would expand the scope of task-based research. From this viewpoint, a task should not be considered as a tool to force learners to follow predicted steps but rather be seen as a pedagogical approach to creating an environment in which they can use and learn L2 in their own autonomous ways. The traditional approach to TBLT has been aimed at predicting how learners will behave by identifying the linear causality of task and performance, while the task-as-affordance perspective illuminates another line of inquiry. The optimal learning environment provides an affordance within which an active learner engages in meaning-making activities, and learning emerges as a consequence of the

"dynamic interplay between learner's performance and the context affordances as perceived by the learner" (Larsen-Freeman, 2009, p. 587).

The importance of learner agency has also been emphasised in some oral-based research. Storch (2008) suggested that elaborate engagement encouraged a deeper level of understanding and was more effective in leading to language learning in collaborative dialogues. However, greater availability of production time and the recursive nature of complex cognitive processes in writing may constitute even more ideal conditions for task repetition to foster deeper linguistic processing as a result of "a gradual complexification of goals being pursued" (Manchón, 2014a, p. 20). Our analysis of the interaction between composing and self-regulatory processes suggests that the writing modality essentially incorporates functions of internal repetition. That is, when students were seriously engaged in the task (as in the case of Yuto), they continuously shifted their attention among various processes (such as idea generation, text construction, revision and evaluation), and the sequence of these processes occurred non-linearly and recursively. Because writing essentially involves internal repetition, it could be beneficial to encourage learners to become aware of this function through self-reflections as well as external repetitions. In other words, repetition is "an integral part of the successful completion of the writing task itself as a function of its inherent complexity and processing demands" (Manchón, 2014a, p. 27).

Together with our previous research, the present study also points out the values of the case-based method, which is useful for understanding how individuals function and change over a significant period of time. To identify whether and how the quality of the writing changed as a result of these iterative operations, we focused on their variability. Different performances and learning outcomes from engaging in task repetition should be regarded as the norm, and not as measurement error. Most of the oral-based research (as presented in Table 1) has been aimed at detecting the effects of repetition by taking a few snapshots, as is typically found in pre-post designs. However, our findings suggested a large degree of intra-individual variability in the students' writing. For example, because Taka's writing was characterised by occasional high values and sudden drops, very different conclusions could be drawn based on which points in time the data were focused on. In terms of inter-individual variability, the differences may not have been clearly observable from comparing the beginning compositions of Yuto and Taka, but distinctive differences emerged over time from the longitudinal observations. In other words, through the process of 'particularisation' rather than 'generalisation', we found insights which could "be adapted to, and provide comparative information to a wide variety of other cases" (van Lier, 2005, p. 198). Clearly, Yuto and Taka are particular cases, but the functions of their self-regulatory processes and emergent patterns in their writing development can probably also be observed within other contexts.

Conclusion

This study explored the ways in which repetition relates to writing development from the perspective of CDS. We questioned research attempting to explain the effects of task repetition through analysis of oral data involving limited repetitions over short periods. This was not only because language learning takes place over longer periods and certain types of tasks are repeatedly used within a classroom context, but also because writing seems to have different pedagogical benefits compared to speaking. We also questioned the reliability on group-level analysis and considered the potential of employing the case-based method to explore the micro-level effects of internal repetition. We believe it is of paramount importance to conduct research that investigates the long-term effects of writing task repetition with sufficient attention being devoted to individual learner differences.

The findings we presented indicated that task repetition should be regarded as dynamic and constructive. Repetition enables learners to explore and challenge their output by making the context and/or the frame stable (Bygate, 2006). Their L2 systems are built up through interactions with the writing task, and this constructive process can be effectively achieved through repetition. Yuto initially engaged in writing to improve his writing fluency, as directed by task instructions in this case study, and then found his own writing goals through repeated engagement in reflecting on his compositions. Our findings also suggest that repetition of the writing task we investigated is likely to push learners into language learning but only when they are actively engaged. We assumed that self-reflective practice would increase their awareness about self-regulatory processes (Zimmerman & Risemberg, 1997) but, as in Taka's case, it is by no means a method that will encourage all students to do so automatically. Thus, it would certainly be necessary to investigate what pedagogical interventions can support this type of learner in future research.

Last but not least, we should emphasise that our aim was not to claim that the traditional approach built on the computational metaphor is less meaningful; rather we expect that they will continue to make significant contributions to the area of TBLT. However, we are concerned that the dominance of only one approach is not a healthy condition and is against epistemological diversity, which "has been a good thing for SLA" (Ortega, 2012, p. 212) and "is here to stay as part of the mainstream of the field" (p. 221). It is hoped that our investigations will offer another epistemological lens and contribute to expanding the value of research on task repetition.

Acknowledgements

We are grateful to Martin Bygate and Kris van den Branden, and anonymous reviewers for providing valuable comments and suggestions on the earlier versions of this chapter.

References

Ahmadian, M. J., & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research*, 15(1), 35–59.

Arevart, S., & Nation, P. (1991). Fluency improvement in a second language. *RELC Journal*, 22(1), 84–94.

Austin, J. T., & Vancouver, J. B. (1996). Goal constructs in psychology: Structure, process, and content. *Psychological Bulletin*, 120(3), 338.

Baba, K., & Nitta, R. (2010). Dynamic effects of task type practice on the Japanese EFL university student's writing: Text analysis with Coh-Metrix. Proceedings of the 23 International Florida Artificial Intelligence Research Society Conference (FLAIRS #2010), 217–222.

Baba, K., & Nitta, R. (2012). Dynamic effects of repeating a timed writing task in two EFL university courses. In P. M. McCarthy, & C. Boonthum-Denecke (Eds.), *Applied Natural Language Processing: Identification, Investigation and Resolution* (pp. 398–413). Hershey, PA: IGI Global.

Baba, K., & Nitta, R. (2014). Phase transitions in the development of writing fluency from a complex dynamic systems perspective. *Language Learning* 64(1), 1–35.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Boekaerts, M., Pintrich, P. R., & Zeidner, M. (2000). *Handbook of self-regulation*. Oxford: Elsevier.

Boud, D., Keogh, R, & Walker, D. (Eds.) (1985). *Reflection: Turning experience into learning*. London: Routledge.

Breen, M. (1987/2009). Learner contributions to task design. In C. Candlin, & E. Murphy (Eds.), *Language learning tasks* (pp. 23–46). London: Prentice Hall. Reprinted in K. van den Branden, M. Bygate, & J. M. Norris (Eds.). (2009). *Task-based language teaching: A reader* (pp. 333–356). Amsterdam: John Benjamins.

Bruner, J. (1983). *Child's talk: Learning to use language*. Oxford: Oxford University Press.

Bygate, M. (1996). Effects of task repetition: Appraising the developing language of learners. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 136–146). Oxford: Macmillan.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks, second language learning, teaching and testing* (pp. 23–48). London: Longman.

Bygate, M. (2006). Areas of research that influence L2 speaking instruction. In E. Uso-Juan & A. Martinez-Flor (Eds.), *Current trends in the development and teaching of the four language skills* (pp. 159–186). Berlin: Mouton de Gruyter.

Bygate, M., & Samuda, V. (2005). Integrative planning through the use of task-repetition. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 37–76). Amsterdam: John Benjamins.

Byrne, D. (2009). Case-based methods: Why we need them; what they are; how to do them. In D. Byrne & C. C. Ragin (Eds.), *The Sage handbook of case-based methods* (pp. 1–13). Los Angeles, CA: Sage.

Byrne, D., & Ragin, C. C. (2009). *The Sage handbook of case-based methods*. Los Angeles, CA: Sage.
Coughlan, P., & Duff, P. (1994). Same task, different activities: Analysis of a SLA task from an Activity Theory perspective. In J. P. Lantolf & G. Appel (Eds.), *Vygotskian approaches to second language research* (pp. 173–194). Westport, CT: Ablex.
Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
de Bot, K., & Larsen-Freeman, D. (2011). Researching second language development from a Dynamic Systems Theory perspective. In M. H. Verspoor, K. de Bot, & W. Lowie (Eds.), *A dynamic approach to second language development: Methods and techniques* (pp. 5–23). Amsterdam: John Benjamins.
de Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition*, 10, 7–21.
de Jong, N., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, 61(2), 533–568.
Dörnyei, Z. (2014). Researching complex dynamic systems: 'Retrodictive qualitative modelling' in the language classroom. *Language Teaching*, 47(1), 80–91.
Ellis, R. (1997). SLA and language pedagogy: An educational perspective. *Studies in Second Language Acquisition*, 19/1, 69–92.
Fukuta, J. (2015). Effects of task repetition on learners' attention orientation in L2 oral production. *Language Teaching Research*, 20(3), 321–340.
Gass, S., Mackey, A., Alvarez–Torres, M. J., & Fernández–García, M. (1999). The effects of task repetition on linguistic output. *Language Learning*, 49, 549–581.
Gibson, J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Miffin.
Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36, 193–202.
Heckhausen, J., & Dweck, C. S. (1998). *Motivation and self-regulation across the life span*. Cambridge: Cambridge University Press.
Kim, Y. (2013). Promoting attention to form through task repetition in a Korean EFL context. In K. McDonough & A. Mackey (Eds.). *Second language interaction in diverse educational contexts* (pp. 3–24). Amsterdam: John Benjamins.
Larsen–Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy and fluency in second language acquisition. *Applied Linguistics*, 30, 579–589.
Larsen–Freeman, D., & Cameron, L. (2008). *Complex systems and applied linguistics*. Oxford: Oxford University Press.
Levelt, W. J. M. (1989). *Speaking: From intention to articulation*: Cambridge, MA: The MIT press.
Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, 4(3), 221–250.
Lynch, T., & Maclean, J. (2001). Effects of the benefits of immediate task repetition on learners' performance. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks, second language learning, teaching and testing* (pp. 141–162). London: Longman.
Mackey, A., Kanganas, A. P., & Oliver, R. (2007). Task familiarity and interactional feedback in child ESL classrooms. *TESOL Quarterly*, 41(2), 285–312.
Manchón, R. M. (2014a). The distinctive nature of task repetition in writing. Implications for theory, research, and pedagogy. *ELIA*, 14, 13–41.
Manchón, R. M. (2014b). The internal dimension of tasks. In H. Byrnes & R. Manchón (Eds.), *Task-based language learning. Insights from and for L2 writing* (pp. 27–52). Amsterdam: John Benjamins.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42, 381–392.

Nakata, Y. (2016). Promoting self-regulation through collaborative work. *Eurosla Yearbook*, 16, 59–84.

Nitta, R., & Baba, K. (2014). Task repetition and L2 writing development: A longitudinal study from a dynamic systems perspective. In H. Byrnes & R. Manchón (Eds.), *Task-based language learning. Insights from and for L2 writing* (pp. 107–136). Amsterdam: John Benjamins.

Nitta, R., & Baba, K. (2015). Self-regulation in the evolution of the ideal L2 self: A complex dynamic systems approach to the L2 Motivational Self System. In Z. Dörnyei, P. MacIntyre, & A. Henry (Eds.), *Motivational dynamics in language learning* (pp. 367–396). Bristol: Multilingual Matters.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492–518.

Ortega, L. (2012). Epistemological diversity and moral ends of research in instructed SLA. *Language Teaching Research*, 16(2), 206–226.

Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics*, 36, 50–72.

Pinter, A. (2007). Some benefits of peer-peer interaction: 10-year-old children practising with a communication task. *Language Teaching Research*, 11(2), 189–207.

Plough, I., & Gass, S. (1993). Interlocutor and task familiarity: Effect on interactional structure. In G. Crookes & S. Gass (Eds.), *Tasks and language learning* (pp. 35–56). Clevedon: Multilingual Matters.

Schunk, D. H., & Zimmerman, B. J. (2008). *Motivation and self-regulated learning: Theory, research, and applications*. New York, NY: Lawrence Erlbaum Associates.

Storch, N. (2008). Metatalk in a pair work activity: Level of engagement and implications for language development. *Language Awareness*, 17, 95–114.

van Geert, P. (2011) The contribution of complex dynamic systems to development. *Child Development Perspectives*, 5(4), 273–278.

van Geert, P., & van Dijk, M. (2002). Focus on variability: New tools to study intra-individual variability in developmental data. *Infant Behavior & Development*, 25, 340–374.

van Lier, L. (2000). From input to affordance: Social-interactive learning from an ecological perspective. In J. P. Lantolf (Ed.). *Sociocultural theory and second language learning* (pp. 245–260). Oxford: Oxford University Press.

van Lier, L. (2004). *The ecology of semiotics of language learning: A sociocultural perspective*. Dordrecht: Kluwer.

van Lier, L. (2005). Case study. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 195–208). Mahwah, NJ: Lawrence Erlbaum Associates.

Weigle, S. C. (2006). Investing in assessment: Designing tests to promote positive washback. In P. K. Matsuda, C. Ortmeier-Hooper, & X. You (Eds.), *The politics of second language writing: In search of the promised land* (pp. 222–244). West Lafayette, IN: Parlor Press.

Zimmerman, B. J. (1998). Developing self-fulfilling cycles of academic regulation: An analysis of exemplary instructional models. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulated learning: From teaching to self- reflective practice* (pp. 1–19). New York, NY: Guilford.

Zimmerman, B. (2002). Becoming a self-regulated learner: An overview. *Theory and Practice*, 41(2), 64–70.

Zimmerman, B. & Risemberg, R. (1997) Becoming a self-regulated writer: A social cognitive perspective. *Contemporary Educational Psychology*, 22, 73–101.

CHAPTER 12

Task repetition or task iteration?

It does make a difference

Diane Larsen-Freeman
University of Michigan & University of Pennsylvania

This chapter recommends a different understanding of task "repetition" from the perspective of Complex Dynamic Systems Theory (CDST). It suggests that a better term than "repetition" is "iteration". While substituting one term for another might seem to be purely an academic exercise, doing so actually has important implications for researchers and teachers. For one, it is an acknowledgement of the common observation that learners' performance on tasks will never be the same from one iteration to the next. In addition, it maintains that useful language pedagogy should build learners' capacity to make meaning beyond their exploiting conventional uses of language to which learners have already been exposed. Further to this point, *iteration* instead of *repetition* is consistent with the CDST position that learning is not limited to reproduction. Through iteration, learners receive practice in adapting their language resources to mutable temporal and spatial contexts, and they construct their own learning paths. Also, the chapter problematizes the notion of transfer of learning, conceiving of it as *transformation* rather than transfer, and it addresses issues of measurement and nonlinearity. Finally, calling it task iteration honors the agency and uniqueness of learners and the construal that they give to the task at hand. As such, it calls for researchers to adopt a more emic perspective.

Introduction

There is much discussion of Task-Based Language Teaching (TBLT) in the professional literature. One reason for its prominence is undoubtedly that the use of tasks promises one solution to the long-standing problem of how to bridge the form-meaning divide; it also endorses the value of interaction in language learning. Well-designed tasks engage students in using the language and not just learning about it. Researchers have promoted the use of tasks, and TBLT has had at least a modest impact on pedagogy so far (Larsen-Freeman, 2015). Its popularity among researchers is evidenced in its having its own conference and in the numerous articles, books (such as this one), and even an entire book series dedicated to it.

DOI 10.1075/tblt.11.12lar

I, myself, have commissioned book authors to design "grammar-essential" tasks (Loschky & Bley-Vroman, 1993) for a student grammar series that I directed (*Grammar Dimensions: Form, Meaning, and Use)* in order to accomplish several objectives: to provide an appropriate context for the target structures, to elicit a student language sample for teachers to assess what students know and can do (and if students demonstrate that they do know, to preserve time for things they don't), and to motivate students who may come to realize that they do not have the requisite knowledge or ability to complete the given task. I recognize that the use of grammar-essential tasks is a circumscribed application of TBLT; nevertheless, there is little doubt in my mind that when tasks are purposefully and judiciously used, they can contribute significantly to one's pedagogical approach to language teaching (Larsen-Freeman & Anderson, 2011).

In order to have TBLT enjoy further approbation, however, it would be good to revisit the foundations on which it rests. TBLT has long been endorsed by researchers whose studies center on the role of input and interaction in second language acquisition (e.g., Long, 1996; Mackey, 1999; Ellis, 2003). More recently, Samuda and Bygate (2008) have found justification for TBLT in the epistemologies of Dewey and Bruner, who support active discovery through problem-solving tasks. Samuda and Bygate (2008) call for using tasks to create "experience-based opportunities for language learning" (p. 36). While I value tasks for encouraging both interaction and experiential learning, I derive my own theoretical support for TBLT from Complex Dynamic Systems Theory (CDST), a theory which deals centrally with change (Larsen-Freeman & Cameron, 2008). The thesis of this chapter is that CDST inspires a different conception of task repetition, a conception which is advanced by replacing "repetition" with "iteration".

CDST's interpretation of learning through task repetition

Many researchers and teachers claim that there are benefits to task repetition, and they offer explanations for why it is beneficial. I do not have the space to review all the research, but here are some of the stated benefits: repetition in task-as-process increases fluency by freeing up learners' attentional resources (e.g., Bygate, 2001; Bygate & Samuda, 2005), within-task repetition promotes automaticity in the use of elicited formulaic speech (Gatbonton & Segalowitz, 2005), and task-type repetition improves writing performance (Nitta & Baba, 2014).

Related to freeing up attentional resources and to promoting automaticity, another prominent position is that task repetition leads to proceduralization, whereby repeated practice automatizes a procedure. De Jong and Perfetti, (2011) list what proceduralization contributes from an information-processing perspective:

> In the context of language we could say that the retrieval speed of words and phrases increases with repeated practice, with large initial gains that gradually diminish to smaller gains. The creation and strengthening of new chunks can lead to the emergence of formulaic sequences. Language use can also lead to the construction of new production rules and the collapsing of production rules into larger ones. Repeated practice is necessary for these collapsed production rules to be able to compete with (and defeat) their "parent" production rules. (p. 538)

Although I am less sure about the collapsing and defeat of production rules, the other putative benefits are attested by many, and the Power Law of Practice (where initial large gains give way to smaller ones) is well-established. Engaging in tasks involves the conventional use of language to some degree, the learning of which is facilitated by frequent encounters with it. However, language is not a closed system, and learning it is not limited to reproduction. Learning a language should build learners' capacity to make meaning beyond exploiting conventional uses to which they have been exposed. Instruction should empower learners with regards to what is, and to what could be in the expression of new meaning. When I say "what could be," I am not speaking of creativity in the Chomskyan sense of applying rules to create and to interpret novel utterances, although this is one example of creativity. Taylor (2012) contrasts this type of creativity with what he calls "innovation":

> Creativity involves the application of the rules of grammar to items selected from the lexicon; creativity therefore remains strictly within the bounds of the language system. Innovation is a matter of going beyond the system. (p. 246)

Congruent with "going beyond the system" in Taylor's sense of innovating is a foundational concept in CDST, that of *emergence*. Emergence means the arising of some novel property or pattern in a system as a result of the interaction of the components that comprise the system, much as a bird flock emerges out of individual birds. Going beyond the extant system, in the case of language, of course, is accomplished by intentional human agents, i.e., its speakers. To this end, Shanker and King (2002) contrast the rule-governed information-processing view of creativity with a CDST position on creativity:

> Whereas the information processing paradigm sees creativity as a property of the language system itself, … [CDST] views creativity as a property of agents' behavior in co-regulated interactions. (p. 608)

In other words, the potential for creative use of language does not reside in the language system, but in the agency of language users/learners engaged in using/learning a system, where the system is open and where novelty and uniqueness are immanent. As Larsen-Freeman (2006) put it, "Language learning is not about conformity to uniformity" (p. 194).

Indeed, CDST problematizes the deep-seated assumption that learning is a matter of reproducing internally what is external to the system (Davis & Simmt, 2003). Instead, DST researchers Smith & Thelen (1993) propose that speakers soft assemble language constructions. The assembly is soft because speakers make real-time adjustments to create meaning and to realize their intentions within a specific context. The constructions are "created and dissolved as tasks and environments change" (Thelen & Bates, 2003, p. 381). The constructions are cobbled together, at times from the different languages in a learner's repertoire, in order to deal with the communicative pressures at hand – a process of bricolage (Levi-Strauss, 1962) or translanguaging (Otheguy, Garcia, & Reid, 2015). Thus, soft assembly is a real-time process which takes into account both the options to make meaning and their constraints that are available to a speaker, the individual's language-using history, the perceived affordances of the context, and the exigencies at the moment.

Furthermore, soft assembly is subject to many other influences (Lowie, 2013), such as the social weighting of the various options (Sumner, Kim, King, & McGowan, 2014), and, relatedly, who one's interlocutors are. There is a clear interlocutor effect on what a learner produces (Gurzynski-Weiss & Baralt, 2014), a finding especially germane to complex systems, which exhibit coordinated behavior between interlocutors (Dale & Spivey, 2006). As Shanker and King (2002) elaborate,

> … mutual understanding is something that *emerges* as both partners converge on some shared feeling, thought, action, intention, and so on. Far from following some predetermined format, such a process is intrinsically creative…Not only must one always be prepared for the possibility that a subject's communicative acts are highly idiosyncratic, but further, one must be alert to the new dimensions of meaning that are co-created in the contexts of ever-changing and evolving interactions. (p. 608)

Over time, some patterns will be preferred, other patterns are more ephemeral and may never recur. They do not (as yet, or perhaps, ever) fit the system that the learner is constructing and/or they are not equally valorized. The point is that learners are not passive recorders of everything in the environment. Indeed, they can even generate their own feedback through anticipation (Spivey, 2007), something called "predictive error" (Jaeger & Ferreira, 2013) or statistical pre-emption (Johnson, Turk-Browne, & Goldberg, 2013; see also Rescorla & Wagner, 1972). This is where the learner expects to find something in the use of language that never materializes. In the absence of the expected feature, the learner has an opportunity to revise his/her expectations. For instance, say a learner of English has encountered the ditransitive verbargument construction "John gave the library some books". As a result, the learner expects the same ditransitive pattern to occur with the verb "donate", i.e., *John donated the library some books", but it never does. The absence of positive evidence gives the learner the information he or she needs to avoid using the ditransitive construction with the

verb "donate". In other words, the controversy over whether or not children receive the negative evidence they need in order to learn their L1 is irrelevant. Learners are capable of generating their own evidence.

CDST is a relational theory. One way that "relationality" is manifest is that a learner's language resources are constructed in relation to what the learner perceives in the environment. Their uniqueness is due to the fact that learners/users perceive and act on the affordances differently (van Lier, 2004; Larsen-Freeman, 2016a). As Mercer (2012) states,

> The learner makes personal sense out of what they encounter and uses affordances in ways that are personally meaningful and relevant…Agency thus emerges from the interaction between resources and contexts and the learners' perceptions and use of them. (p. 43)

Another way of expressing a similar sentiment is to say "[b]ehavior is not something that a person 'has' – instead, it emerges from interactions between the individual and his or her contexts" (Rose, Rouhani, & Fischer, 2013, p. 153).[1]

Returning now to task repetition, the point is that engaging in a task more than once can be a helpful classroom practice, but its value is not limited to the mapping of conventional forms and meanings or to the proceduralization of existing knowledge. The practice is helpful because it gives learners opportunities to employ heuristics (e.g., exaption) to construct patterns in their own understanding and deployment of their language resources.[2] Proceduralization is practicing what is. Constructing new patterns helps learners go beyond the input. It is building capacity, rather than competence (Widdowson, 1983; Samuda, 2001).

Shanker and King (2002, p. 616) make much the same point in writing about the ape Kanzi's acquisition of language. They contrast Deacon's information-processing position with that of ape researcher Savage-Rumbaugh's dynamic systems approach. Deacon (1997) sought to explain how Kanzi was able to acquire a pre-existing structure or grammar. Savage-Rumbaugh's dynamic systems approach (e.g., Savage-Rumbaugh & Lewin, 1994), on the other hand, aimed to understand how Kanzi's

1. To extend this position even further, in their discussion of contemporary views of the brain as a complex, self-organizing system with nonlinear dynamics, Fazelpour and Thompson (2015) point out the important shift in thinking that has taken place, which attributes the arising of mental phenomena to the brain's self-organized and spontaneous, pattern-generating activity, and not simply from stimulus-driven processing.

2. It may seem that I am describing constructivism; however, Piagetians have trouble with explaining the creation of novelty, which CDST attributes to the fundamental property of *emergence*. In this way, van Geert (2017, p. 34) concludes "The modern theory of complex dynamical systems adds an important foundation to the original Piagetian scheme of explanation…the constructivist program needs to merge with that of complex dynamical systems."

verbal skills emerged in the context of, and was shaped by, the co-regulated interactions in which he engaged.

All this brings me to my thesis about the need to recruit new terms for new ways of thinking. Informed by CDST, I recommend the word *iteration* (Larsen-Freeman, 2012) to replace "repetition". Repetition suggests a self-sameness that is not warranted. As applied linguists know, how people use language makes a difference. Therefore, I suggest using "iteration" to make explicit the claim that iterating a performance on a task brings about a change in capacity. As Doll (1993) puts it, repetition "is designed to improve set performance. Its frame is closed…Iteration[3] aims at developing [capacity]…Its frame is open" (p. 256).

Applications to research and teaching

I will now turn to consider how the views that I have been outlining shape notions of TBLT and its use in research and teaching.

Iteration, not repetition

Clearly, when tasks-as-workplan are repeated, this does not mean that there is repetition of task-as-process (Ellis, 2003).[4] What a CDST perspective adds is that any successive task-as-process should probably be better thought of as "constructive iteration", not repetition.[5] What gets constructed is the complex system of the learner's language resources. As complex systems iterate over time, their structure is built up, both within the individual and within the community. Each iteration reuses some of the patterns generated in the previous procedure(s) of soft assembly, always starting at a different point. Van Geert and Verspoor (2015) state that in a dynamic system, "[t]he iterative nature of the processes involved is central to the notion of development, where the next 'state' of development is a function of the preceding state and a condition for the next state" (p. 538).

---

3. Doll calls it "recursion." He also uses "competence" rather than "capacity". However, he does not mean "competence" in the universal grammar sense.

4. Even saying task-as-workplan can be repeated is not entirely true. We know that speakers do not say the same sound the same way twice (Milroy & Milroy, 1999), and evidence at present suggests that our perceptual systems are extremely mutable as well (Eisner & McQueen, 2006).

5. Bygate (2006, p. 172) calls it "constructive repetition" to distinguish it from verbatim or identical repetition. This is a step in the right direction, but I think a new term is called for in order to avoid confusion.

Therefore, calling it task iteration, rather than repetition, avoids giving the misleading impression that using the same task more than once will elicit the same response each time. To expect it to do so is to ignore the history of the individual and shared language resources at that moment and the unique emic perspective of those engaged in a task. Learners perceive and act on affordances differently. As one example of this, Kim and Tracy-Ventura's (2013) task-based interaction research has shown how learners' familiarity with a task affects the number of interactional features, such as clarification requests and language-related episodes, that occur.

CDST also encourages viewing phenomena at different levels of scale. When researchers do this, different effects of repetition present themselves. For instance, fMRI images reveal that repetition has a differential effect in the brain depending on whether what is being repeated is novel or known (Weber, Christiansen, Petersson, Indefrey, & Hagoort 2016). Such findings may explain Davis and Sumara's observation (as reported in Ahmadian & Tavakoli, 2011) that "… very similar systems under virtually identical circumstances and subjected to virtually identical stimuli can respond in dramatically different ways" (pp. 129–130). In this respect, the same authors add that "(t)he *same* system …can and will respond very differently to sets of conditions that appear identical (Davis & Sumara 2006, 18; original emphasis)." In addition, there are multiple pathways by which the system can evolve and "the same 'cause' can, in specific circumstances, produce different effects" (Urry, 2005, p. 4). This is termed the principle of equifinality, and it is characteristic of an open system such as language, as opposed to a closed one where a direct cause and effect relationship exists between the initial state and final state of the system.

That is not to say that the same task-as-workplan cannot be used more than once in a classroom. In fact, from a CDST standpoint, using a task-as-workplan more than once can drive learning. Indeed, when it comes to language learning, revisiting similar territory again and again is essential (Larsen-Freeman, 2013b). Perhaps procedural repetition, as opposed to content repetition, is a helpful way to accomplish this (Kim & Tracy-Ventura, 2013), provided it is possible to determine that one procedure is not radically different from another.

In sum, I am maintaining that each time the same task is used, the learners' experience of it will be different, in part because learners will orient to it differently. Depending on what they are working on, they will extract a different lesson. Besides, from a CDST perspective, learning is not the taking in of different linguistic forms in an aggregative manner; it is changing the system (Feldman, 2006). This happens most efficaciously when learners are engaged in exploiting the meaning potential of the language with each iteration of a task. Although it may seem that I am splitting hairs, I believe that the difference between exact repetition and meaningful iteration is a non-trivial one.

Adaptation

In environments where natural communication takes place, interlocutors have to deal with unpredictability, interruptions, and distractions. In such cases they also have to notice changes in the environment and respond to them rapidly (Segalowitz & Tromfimovich, 2012). Thus, they have to learn to deal with changes as they occur in real time by adapting to an interlocutor or to different task demands. Of course, learners can do this already in a language with which they are familiar, but when it is a situation involving a language for which they have lesser facility, the opportunity to encounter just a modest change will assist them in learning to adapt their language resources to a context, which is always in transition.

Iterating tasks can play a role in helping students learn how to adapt their language resources to a mutable temporal and spatial context by changing the original task-as-workplan in subsequent iterations. Even a slight alteration (which is what biologists tell us optimizes adaptation) in the plan provides an opportunity for learners to make different choices, while adapting to a different context (Larsen-Freeman, 2013b), thus changing their language resources.

Of course, most teachers will naturally build from one task or activity to the next. What may be different in what I am recommending is for teachers to teach learners to adapt by implementing only a minimal change in the next iteration of the task. Take for example, Samuda's (2001) 'Things-in-Pockets' task where students are asked to speculate about the identity of a person when shown the contents of this person's pockets. An iteration of this task might involve replacing two of the items in the person's pockets with two new items. The new items, if carefully chosen, could challenge students' conclusion from the previous iteration and lead to a new round of speculation about the person's identity. Alternatively, students might be asked to engage in the same task with a different partner, or, changing the conditions, be given less time on each iteration to accomplish the task as in a 4-3-2-type activity (Maurice, 1983). Slight alterations in task-as-workplan teaches students to take their language resources and mould them to a changing situation, which makes different demands.

Different learning paths

Whatever modifications to a task a teacher/researcher makes, it is important to remember that students not only start from different points when they begin to engage with a task, they also make their own way. In this respect, Skehan (2009) noted that different task participants seemed to gravitate to different aspects of form. Along the same lines, Kim, Payant, and Pearson (2015) found that learners with lower working memory capacity may not benefit as much from complex tasks as those with higher

capacity. Significant for the present discussion, there is also evidence that not all learners benefit from repeating a task-as-workplan (Ahmadian, 2013) or benefit equally.

Finding such differences among individuals is why many years ago I abandoned my search for an SLA index of development. I had sought a measure that would work for both individuals and groups on various products of learner performance. For instance, the average length of error-free T-units in writing worked well in discriminating among groups of learners, but did not necessarily reflect development for individuals. Of course, it has long been known that group data may describe a process, or a functional relation, which has no validity for any individual (Sidman,1960). Extapolating findings from sample-based research to individual process statements is often demonstrably incorrect (van Geert, 2011).[6] Even monozygotic twins, with similar exposure to language and similar experiences, follow different developmental paths (Verspoor, Schmid, & Xu, 2012; Chan, Verspoor, & Vahtrick, 2015).

Most findings on the effectiveness of TBLT using quantitative methods are based on statistical testing using group means. While performance by groups provides one measure of the success of an intervention, CDST encourages assessing performance at different levels of granularity. Dropping down one level to the level of the individual, a great deal of variability can be seen following the implementation of an intervention. To give one example, Nguyen's study of participants' learning of formulaic utterances through spot-the-difference and collaborative gap-fill tasks reveals individual differences to such a degree that certain learners in the control group outperformed the experimental participants, and certain experimental participants did not seem to benefit at all from the instruction that they had received (Nguyen & Larsen-Freeman, 2017).

Moreover, comparing groups in an experiment where the independent variable is a pedagogical intervention, such as a task, has often entailed null hypothesis statistical testing (NHST), which has come under increasing criticism. NHST has found less favor because the chance of obtaining statistically significant differences can be increased by, among other things, simply increasing the sample size (Plonsky, 2014).

> If we want to make statements about *general trends* of factors affecting products of development, then group generalizations are useful, especially when applying

6. However, this fact does not rob us of the ability to generalize (Larsen-Freeman, 2017a). For instance, van Geert (2011) explains that a single case study can be generalizable, depending on how it links to a particular theory:

> For students of language development, single case studies have a direct bearing on the underlying theory, and only an indirect one on the population of language learners. In summary, a truly general theory of developmental processes is one that can be "individualized" – it can generate theory-based descriptions of individual trajectories in a nontrivial sense. (p. 276)

> Bayesian statistics and provided the focus is on power and effect sizes rather than on NHST. (Lowie, 2015)

However, the one thing that is known for certain is that the average is not good enough if the goal is to help all learners succeed. Labelling certain learners 'outliers' because they do not fit a normal curve or merely reporting standard deviations is to sacrifice these learners, overlook what makes them unique, and risks not taking advantage of a source of insight into the developmental process that learners' task performance could provide.[7]

Transfer

Another issue raised by adopting a CDST perspective on task repetition is to question whether "transfer" is the appropriate word to characterize what students do when their learning from one task is subsequently applied to another. This is an important area of investigation since learning transfer, or the ability for students to apply what they have learned in school to another setting, is the whole premise upon which formal education is based.

So far the evidence from task repetition has been mixed. For instance, Bygate (2001) reports that learners did not show the transfer of any benefit from generic task practice to performance on new tasks. He suggests that the absence of generic practice effects is disappointing and warrants further investigation. Ellis (2009) also remarks of his study that "[g]iven that the effects of repeating a task did not transfer to the performance of a new task, there is no clear evidence that task rehearsal assists acquisition" (p. 477).

On the other hand, de Jong and Perfetti (2011) did report finding transfer, but only after tasks preceding the opportunity for transfer were repeated three times. Ahmadian (2012) notes that the empirical evidence regarding the extension of the effects of task repetition to a new task is still relatively scarce. He adds:

> Arguably, if the benefits of repeating the same task do not transfer when a new task is performed, one may conclude that task repetition does not assist L2 acquisition and that the effects of task repetition are limited to immediate L2 performance (Ellis 2009; Larsen-Freeman 2009). (p. 381)

This observation supports Gass et al. (1999) who found that task repetition resulted in improved student performance, but the improvements did not generalize to a new context.

7. Even traditional research on individual differences is conducted on groups of learners.

As I have written, though, one issue raised by adopting a CDST perspective on task repetition is to question whether 'transfer' is the appropriate term for what is intended. Of course, because of its importance, there has been a long tradition of educational research on learning transfer (Larsen-Freeman, 2013a). However, 'transfer' evokes a transportation metaphor, which suggests that what is learned in one situation is carried over to another. CDST categorically rejects this interpretation and concurs with Lobato's (2006) observation that the term 'transfer' suggests that "the tasks or situations across which transfer occurs remain unchanged during the transfer and that the 'transferor' reproduces existing relations between fixed tasks" (Lobato, 2006, p. 444). In other words, explanations for transfer are based on the psychological invariance of mental representations and action schemata.

According to CDST, there is no such thing as invariant mental representations. Spivey (2007) warns that we need to drop:

> …the assumption of stable symbolic internal representations (holdovers from an information-processing perspective on cognitive processes)…continuing on to a fully ecological dynamical account of perception, cognition, and action that connects the brain, body, and world. (p. 332)

Instead of manipulating invariant mental representations, CDST posits that learners do not simply unload a prior solution. Instead, they craft it on the spot, adjusting and adapting their prior knowledge in the process (Carraher & Schliemann, 2002)."[8]

From a CDST point of view, the crucial point is that linguistic knowledge is not *given or transferred* but adaptively *achieved* by the individual in the environment (Leather & van Dam, 2003). Such a characterization implies that *transformation* would be a better word than 'transfer'. Learners transform their knowledge; they do not merely implement knowledge in the form in which it was delivered through instruction (Larsen-Freeman, 2013a). In any case, even for transfer, as it has traditionally been depicted, to take place, it is the learners who have to recognize the similarity of the two tasks, and this does not always happen (Gick & Holyoak, 1980). Helping students to see the similarity is a prerequisite to transformation.

Measurement

Another difficulty in studying the effect of task repetition is determining what to measure when students engage in a task. First, there is no way to know how psychologically

8. But, it is not only knowledge or mental representations that are "transferred." When one communicates, desires, intentions, thoughts, wishes, and emotions are also conveyed (Shanker & King, 2002).

real linguistic units are for learners (Tomasello, 2000). We should not therefore limit ourselves to analyzing conventional linguistic units of one type, assuming that learners' units are isomorphic with those of linguists. Second, there is the matter of learners' focusing their attention on aspects of task performance in ways that are not intended by researchers (Kim, Payant, & Pearson, 2015). Third, more developmentally sensitive measures of task performance than those of complexity, accuracy, and fluency (CAF), the qualities most often measured by task-based researchers, are needed (Lambert & Kormos, 2014), including how task participants make use of other semiotic resources, such as gestural ones.

Norris and Ortega (2009) also mention the pattern that has been identified many times in the CDST literature, namely the interconnectedness and reciprocal nature of subsystems during the dynamic process of change. If we examine the dimensions one by one, we miss their complex interaction and the fact that the pattern of their interaction changes with time as well. Indeed, CDST would encourage us to look at how the whole configuration changes over time (Shanker & King, 2002) and to attempt to understand the changes from the perspective of the learners involved.

Byrnes (2007) identifies what has been a further limitation: Task processing is primarily interpreted in individual, psycholinguistic terms. Kim and Tracy-Ventura (2013) note this as well:

> It may also be the case that three repetitions of collaborative tasks may not have affected fluency as measured on monologic speaking tests, but it may have improved interactional competence which is not assessed by traditional psycholinguistic fluency measures… (p. 838)

Locating language in the interstices between people and context, rather than only within the tasks or only within individuals themselves, requires a different approach to language from that of traditional CAF measures. In particular, it suggests that we need to employ more socially oriented measures of development, such as the level of participation that a learner evinces, for instance, or the type of discourse practices in which he or she engages (Larsen-Freeman, 2009).

The need to accommodate individuals extends to teachers, and researchers, too, for that matter. Both teachers and researchers are agents in the classroom ecology. For example, Kim and Tracy-Ventura (2013) reflect on their study of task repetition in the classroom:

> In terms of the ecological validity of the current study, the classroom-based nature of the study was positive, yet the role of the instructor in [sic] classroom had to be minimized to control for any instructor effects. Future studies would benefit from taking the instructor's role into consideration when addressing the role of task repetition during collaborative tasks. (p. 839)

In addition to accommodating agents more generally, we should not overlook the specific intentions, goals, drives, interests, values, evaluations, etc. (cf. van Geert, 2017) of the particular individuals involved in completing the task. For instance, van de Guchte et al. (2015, p. 2) report that the participants in their pilot study found the prospect of task repetition "boring", and some time ago, Plough and Gass (1993) suggested that if tasks are repeatedly used in class, learners can become disinterested due to their lack of novelty. It is likely that when students find tasks boring or uninteresting, they participate less fully than how they would if they found them engaging.

Finally, it is important to honor the context-dependence of complex systems. Norris and Ortega call for accounts that are sensitive to the local ecology – in other words, they recognize that CAF is conditioned by the learning context. As Norris and Ortega state (2009), in a position consistent with CDST, "… it is illusory to think that what we are measuring in CAF is some kind of universal construct that can be applied across all possible learners and contexts" (p. 575).

Nonlinearity

To complicate matters further, from a CDST perspective, while there may be no immediate evidence that task iteration assists learning, this does not mean that it does not. It is certainly plausible that there is a nonlinear relationship between a learner's iterative performance and that same learner's showing improvement from a target-language perspective. In fact, the effect of iterating a task may not occur within the timeframe of a research design, even one with a delayed post-test (Koopmans, 2014). In other words, the benefits of task iteration may not be immediately manifest. At other times it may seem that a learner has made sudden leaps, similar to the dynamic of punctuated equilibrium that Gould (2007) has discussed with respect to evolution.

Connecting timescales (e.g., real time and over time) is recognized to be very important in dynamic systems research (Monaghan & Christiansen, 2010), as it is in the study of second language development more generally. A CDST researcher would say that what is needed is a way to describe and explain how short-term actions and events affect long-term development. What is problematic when it comes to task effects is that there may not be a linear causal connection between performance on a task and learning outcomes. For example, Wolfe-Quintero et al. (1998) report a nonlinear relationship between syntactic subordination in writing and development, with increases in subordination at lower and intermediate levels but decreases at the highest levels.

Although a learner's language production may not look qualitatively different, it is possible that our view of development is obscured because our instruments are too blunt or we are not looking in the right places. For instance, there may be no qualitative phase shift in the system, but there may well be one of integration (Samuda &

Bygate, 2008) or consolidation. In any event, a complex systems view suggests that the potential for development is always present, and that there is no finite state at which it ceases (Larsen-Freeman, 2006).

Conclusion

I began this chapter by acknowledging that CDST inspired my own appreciation of the use of tasks in language teaching, so I should make it clear that my purpose in writing this chapter is not to denigrate the use of tasks, nor using them more than once. Instead, I have suggested that CDST offers a different perspective on task repetition and this leads me to recommend using a different term to reflect this perspective. I see task *iteration* as a better way of reflecting the fact that there is never exact replication when a task is used repeatedly. Moreover, *iteration* does not preclude the emergence of novelty. I have also suggested that slightly altering and iterating tasks-as-workplans (one way of accomplishing this might be through procedural repetition) can help with *adaptation* – teaching students to adapt their language resources to ever-changing situations. I have also proposed substituting *transformation* for 'transfer' to capture the idea that learners transform what they have learned; they do not merely transport their knowledge intact.

CDST proposes that learning is not a process of internalizing an external reality. Rather, through iterated opportunities to make meaning in specific contexts, learners perceive and act on their own affordances, which assist them in constructing their language resources. A traditional way of looking at improved learner performance on repeated tasks might be to say that a learner is increasingly able to perform his or her competence, whereas the approach that I am putting forward in this chapter argues for a dynamic view where there is no distinction between performance and competence. With each use of language, a learner's language resources change. And in this way, capacity is built.

The question which remains for researchers to answer is whether performance on iterative tasks can be transformed in a way that contributes to this change in terms of long-term development. In other words, can we find a way to connect the timescales between immediate use and over time gains (i.e., the same question that all SLA researchers must address)? In order to answer this question, we will need to broaden how we look at and measure language use beyond CAF to include more socially relevant criteria, and we will need to take into account not only linear cause-effect learning of conventional language use, but also the nonlinearity of the learning process and innovative language use. Further, if such research is to be helpful in elucidating language development, it must be carried out over time. Finally, we must not fail to appreciate learners as the unique individuals that they are even though for different

purposes, such as language policy (van Geert & Steenbeek, 2014), we might seek to identify 'typical' learners by generalizing at the level of the group. Perhaps attending to learners as individuals will ultimately call for researchers to adopt a more "hermeneutic stance, looking at what their behaviors mean to the participants as they are involved in co-regulated interactions" (Shanker & King, 2002, p. 609). Doing so will involve more than conducting and comparing case studies (Larsen-Freeman, 2017b). Learning how to study task iteration from the perspective of the learners themselves is a worthy goal indeed (Larsen-Freeman, 2016b).

References

Ahmadian, M. J. (2012). Task repetition in ELT. *ELT Journal*, *66*(3), 380–382. https://doi.org/10.1093/elt/ccs020

Ahmadian, M. J. (2013). Working memory and task repetition in second language oral production. *Asian Journal of English Language Teaching*, *23*, 37–55.

Ahmadian, M. J., & Tavakoli, M. (2011). Exploring the utility of action research to investigate second-language classrooms as complex systems. *Educational Action Research*, *19*(2), 121–136. https://doi.org/10.1080/09650792.2011.569160

Bygate, M. (1999), Task as context for the framing, reframing, and unframing of language. *System*, *27*, 33–48.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 23–48). Harlow: Pearson Longman.

Bygate, M. (2006). Areas of research that influence L2 speaking instruction. In E. Usó-Juan & A. Martínez-Flor (Eds.), *Current trends in the development and teaching of the four skills* (pp. 159–186.) Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110197778.3.159

Bygate, M., & Samuda, V. (2005). Integrative planning through the use of task-repetition. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 37–74). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.11.05byg

Byrnes, H. (2007). Second and foreign language pedagogy. Paper presented at the 30th anniversary of American Association for Applied Linguistics. Retrieved from: <www.ibrarian.net/…/Second_and_Foreign_Language_Pedagogy.pdf> (1 June, 2016).

Carraher, D., & Schliemann, A. D. (2002). The transfer dilemma. *Journal of the Learning Sciences*, *11*, 1–24. https://doi.org/10.1207/S15327809JLS1101_1

Chan, H. P., Verspoor, M., & Vahtrick, L. (2015). Dynamic development in speaking versus writing in identical twins. *Language Learning*, *65*(2), 298–325. https://doi.org/10.1111/lang.12107

Dale, R., & Spivey, M. J. (2006). Unraveling the dyad: Using recurrence analysis to explore patterns of syntactic coordination between children and caregivers in conversation. *Language Learning*, *56*(3), 391–430. https://doi.org/10.1111/j.1467-9922.2006.00372.x

Davis, B., & Simmt, E. (2003). Understanding learning systems: Mathematics education and complexity science. *Journal for Research in Mathematics Education*, *34*(2), 137–167. https://doi.org/10.2307/30034903

Davis, B., & Sumara, D. (2006). *Complexity and education: Inquiries into learning, teaching, and research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Deacon, T. W. (1997) *The symbolic species: The coevolution of language and human brain*. New York, NY: W. W. Norton.

de Jong, N., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, *61*(2), 533–568. https://doi.org/10.1111/j.1467-9922.2010.00620.x

Doll, W. E.Jr. (1993). *A post-modern perspective on curriculum*. New York, NY: Teachers College Press.

Eisner, F., & Macqueen, J. M. (2006). Perceptual learning in speech: Stability over time (L). *Journal of the Acoustic Society of America*, *119*(4), 1950–1953. https://doi.org/10.1121/1.2178721

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity and accuracy in L2 oral production. *Applied Linguistics*, *30*, 474–509. https://doi.org/10.1093/applin/amp042

Fazelpour, S., & Thompson, E. (2015). The Kantian brain: Brain dynamics from a neurophenomenological perspective. *Current Opinion in Neurobiology*, *31*, 223–229. https://doi.org/10.1016/j.conb.2014.12.006

Feldman, J. (2006). *From molecule to metaphor: A neural theory of language*. Cambridge, MA: The MIT Press.

Gass, S., Mackey, A., Fernandez, M., & Alvarez-Torres, M. (1999). The effects of task repetition on linguistic output. *Language Learning*, *49*(4), 549–80. https://doi.org/10.1111/0023-8333.00102

Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *Canadian Modern Language Review*, *61*(3), 325–353. https://doi.org/10.3138/cmlr.61.3.325

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*, 306–355. https://doi.org/10.1016/0010-0285(80)90013-4

Gould, S. J. (2007). *Punctuated equilibrium*. Cambridge, MA: Harvard University Press.

Gurzynski-Weiss, L., & Baralt, M. (2014). Exploring learner perception and use of task-based interactional feedback in face-to-face and CMC modes. *Studies in Second Language Acquisition*, *36*(1), 1–37. https://doi.org/10.1017/S0272263113000363

Jaeger, F. T., & Ferreira, V. (2013). Seeking predictions from a predictive framework. *Behavioral and Brain Sciences*, *36*(4), 359–360. https://doi.org/10.1017/S0140525X12002762

Johnson, M., Turk-Browne, N. B., & Goldberg, A. (2013). Prediction plays a key role in language development as well as processing. *Behavioral and Brain Sciences*, *36*(4), 360–361. https://doi.org/10.1017/S0140525X12002609

Kim, Y., & Ventura-Tracy, N. (2013). The role of task repetition in L2 performance development: What needs to be repeated during task-based interaction? *System*, *41*, 829–841. https://doi.org/10.1016/j.system.2013.08.005

Kim, Y., Payant, C., & Pearson, P. (2015). The intersection of task-based interaction, task complexity, and working memory. *Studies in Second Language Acquisition*, *37*(3), 549–581. https://doi.org/10.1017/S0272263114000618

Koopmans, M. (2014). Change, self-organization and the search for causality in educational research and practice. *Complicity: An International Journal of Complexity and Education*, *11*(1), 20–39.

Lambert, C., & Kormos, J. (2014). Complexity, accuracy, and fluency in task-based l2 research: Toward more developmentally based measures of second language acquisition. *Applied Linguistics*, *35*(5), 607–614. https://doi.org/10.1093/applin/amu047

Larsen-Freeman, D. (2006). Second language acquisition and the issue of fossilization: There is no end and there is no state. In Z-H. Han & T. Odlin (Eds.), *Studies of fossilization in second language acquisition* (pp. 189–200). Clevedon: Multilingual Matters.

Larsen-Freeman D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, *30*(4), 579–589. https://doi.org/10.1093/applin/amp043

Larsen-Freeman, D. (2012). On the roles of repetition in language teaching and learning. *Applied Linguistics Review*, *3*(2), 195–210. https://doi.org/10.1515/applirev-2012-0009

Larsen-Freeman, D. (2013a). Transfer of learning transformed. *Language Learning*, *63*(Special issue), 107–129. https://doi.org/10.1111/j.1467-9922.2012.00740.x

Larsen-Freeman, D. (2013b). Complex, dynamic systems and technemes. In J. Arnold Morgan & T. Murphey (Eds.), *Meaningful action: Earl Stevick's influence on language teaching* (pp. 190–201). Cambridge: Cambridge University Press.

Larsen-Freeman, D. (2015). Research into practice: Grammar learning and teaching. *Language Teaching*, *48*(2), 263–280. https://doi.org/10.1017/S0261444814000408

Larsen-Freeman, D. (2016a). Shifting metaphors: From computer input to ecological affordances to adaptation. A plenary address delivered at the 50th International IATEFL Conference. Birmingham, 15 April. in *Proceedings from the IATEFL 50th Anniversary Conference, Birmingham* (pp.10–19). Kent: IATEFL.

Larsen-Freeman, D. (2016b). The psychology of language learning and "the science of the individual." Paper presented at PLL2, Jyväskylä, Finland, 24 August.

Larsen-Freeman, D. (2017a). Complexity theory: The lessons continue. In L. Ortega & Z-H. Han (Eds.), *Complexity theory and language development: In celebration of Diane Larsen-Freeman* (pp. 11–50). Amsterdam: John Benjamins.

Larsen-Freeman, D. (2017b). On particularizing second language development. A paper presented at the Second Language Research Forum, Ohio State University, 13 October.

Larsen-Freeman, D., & Anderson, M. (2011). *Techniques and principles in language teaching*. Oxford: Oxford University Press.

Larsen-Freeman, D., & Cameron, L. (2008). *Complex systems and applied linguistics*. Oxford: Oxford University Press.

Leather, J., & van Dam, J. (2003). *Towards an ecology of language acquisition*. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-017-0341-3

Levi-Strauss, C. (1962). *La pensée sauvage (The savage mind, English translation,1966)*. Chicago IL: The University of Chicago Press.

Lobato, J. (2006). Alternative perspectives on the transfer of learning: History, issues, and challenges for future research. *Journal of the Learning Sciences, 15*, 431–449. https://doi.org/10.1207/s15327809jls1504_1

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). San Diego, CA: Academic Press.

Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based methodology. In G. Crookes & S. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 123–167). Clevedon: Multilingual Matters.

Lowie, W. (2013). The CEFR and the dynamics of second language learning: Trends and challenges. *Language Learning in Higher Education*, *2*(1), 17–34. https://doi.org/10.1515/cercles-2012-0002

Lowie, W. (2015). Replication in Complex Dynamic Systems approaches to second language development. A paper presented at the American Association of Applied Linguistics/Canadian Association of Applied Linguistics joint conference, Toronto, Canada, 21 March.

Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition*, *21*, 557 – 587. https://doi.org/10.1017/S0272263199004027

Maurice, K. (1983). The fluency workshop. *TESOL Newsletter*, *17*(4), 29.

Mercer, S. (2012). The complexity of learner agency. *Apples – Journal of Applied Language Studies*, *6*(2), 41–59.

Milroy, J., & Milroy, L. (1999). *Authority in language* (3rd ed.). New York, NY: Routledge.

Monaghan, P., & Christiansen, M. H. (2010). Words in puddles of sound: Modelling psycholinguistic effects in speech segmentation. *Journal of Child Language*, *37*, 545–564. https://doi.org/10.1017/S0305000909990511

Nitta, R., & Baba, K. (2014). Task repetition and L2 writing development. In H. Byrnes & R. Manchón (Eds.), *Task-based language learning: Insights from and form L2 writing* (pp. 107–136). Amsterdam: John Benjamins.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA. *Applied Linguistics*, *30*, 555–578. https://doi.org/10.1093/applin/amp044

Nguyen, H., & Larsen-Freeman, D. (2017). Using tasks to teach formulaic sequences: Interindividual and intraindividual variation. In M. J. Ahmadian & M. P. G. Mayo (Eds.), *Recent trends in task-based learning and teaching* (pp. 167–193). Berlin: De Gruyter Mouton.

Otheguy, R., García, O., & Reid, W. (2015). Clarifying translanguaging and deconstructing named languages: A perspective from linguistics. *Applied Linguistics Review*, *6*(3), 281–307. https://doi.org/10.1515/applirev-2015-0014

Plonsky, L. (2014). Study quality in quantitative L2 research (1990–2010): A methodological synthesis and call for reform. *The Modern Language Journal*, *98*(1), 450–470. https://doi.org/10.1111/j.1540-4781.2014.12058.x

Plough, I., & Gass, S. (1993). Interlocutor and task familiarity: Effects on interactional structure. In G. Crookes & S. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 35–56). Clevedon: Multilingual Matters.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning, II: Current research and theory* (pp. 64–99). New York, NY: Appleton-Century-Crofts.

Rose, L. T., Rouhani, P., & Fischer, K. W. (2013). The science of the individual. *Mind, Brain and Education*, *7*(3), 152–158. https://doi.org/10.1111/mbe.12021

Samuda, V. (2001) Guiding relationships between form and meaning during task performance: The role of the teacher. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 119–149). Harlow: Pearson Longman.

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. Houndmills: Palgrave MacMillan.

Savage-Rumbaugh, S., & Lewin, R. (1994). *Kanzi: The ape at the brink of the human mind*. New York, NY: John Wiley & Sons.

Segalowitz, N., & Trofimovich, P. (2012). Second language processing. In S. Gass & A. Mackey (Eds.), *Handbook of second language acquisition* (pp. 179–192). London: Routledge.

Shanker, S. J., & King, B. J. (2002). The emergence of a new paradigm in ape language research. *Behavioral and Brain Sciences*, *25*, 605–656.

Sidman, M. (1960). *Tactics of scientific research*. New York, NY: Basic Books.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510–532. https://doi.org/10.1093/applin/amp047

Smith, L., & Thelen, E. (Eds). (1993). *A dynamic systems approach to development*. Cambridge, MA: The MIT Press.

Spivey, M. (2007). *The continuity of mind*. Oxford: Oxford University Press.

Sumner, M., Kim, S. K., King, E., & McGowan, K. (2014). The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology*, *4*(1015), 1–13.

Taylor, J. R. 2012. *The mental corpus: How language is represented in the mind*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199290802.001.0001

Thelen, E., & Bates, E. (2003). Connectionism and dynamic systems: Are they really different? *Developmental Science*, *6*, 378–391. https://doi.org/10.1111/1467-7687.00294

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, *74*, 209–253. https://doi.org/10.1016/S0010-0277(99)00069-4

Urry, J. (2005). The complexity turn. *Theory, Culture & Society*, *22*(5), 1–14. https://doi.org/10.1177/0263276405057188

van de Guchte, M., Braaksma, M., Rijlaarsdam, G., & Bimmel, P. (2015). Focus on form through task repetition in TBLT. *Language Teaching Research*, *20*(3), 1–21.

van Geert, P. (2011). The contribution of complex dynamic systems to development. *Child Development Perspectives*, *5*(4), 273–278. https://doi.org/10.1111/j.1750-8606.2011.00197.x

van Geert, P. (2017). Constructivist theories. In Hopkins, B. (Ed.). *The Cambridge encyclopedia of child development* (2nd ed.) (pp. 19–34). Cambridge (MA): Cambridge University Press.

van Geert, P., & Steenbeek, H. (2014). The good, the bad and the ugly? The dynamic interplay between educational practice, policy and research. *Complicity*, *11*(2), 22–39.

van Geert, P., & Verspoor, M. (2015). Dynamic systems and language development. In B. MacWhinney & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 537–556). Malden, MA: Wiley-Blackwell.

van Lier, L. 2004. *The ecology and semiotics of language learning: A sociocultural perspective*. Dordrecht: Kluwer. https://doi.org/10.1007/1-4020-7912-5

Verspoor, M., Schmid, M. S., & Xu, W. (2012). A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing*, *21*, 239–263. https://doi.org/10.1016/j.jslw.2012.03.007

Weber, K., Christiansen, M. H., Petersson, K. M., Indefrey, P., & Hagoort, P. (2016). fMRI syntactic and lexical repetition effects reveal the initial stages of learning a new language. *The Journal of Neuroscience*, *36*(26), 6872–6880.

Widdowson, H. G. (1983). Competence and capacity in language learning. In M. Clarke & J. Handscombe (Eds.), *On TESOL '82: Pacific perspectives on language learning and teaching*, (pp. 97–106). Washington, DC: TESOL.

Wolfe-Quintero K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity*. Honolulu, HI: Second Language Teaching & -Curriculum Center, University of Hawaii Press.

External reviewers

Chapters for publication in the book series *Task-Based Language Teaching* are all sent for anonymous external review. I am very grateful to the following colleagues for their readiness to give their time and expertise to review the chapters in this volume:

Patsy Duff
(University of British Columbia),
Pauline Foster
(St Mary's University, London),
Jan Hulstijn
(Emeritus Professor, University of Amsterdam),
Folkert Kuiken
(University of Amsterdam),
Rosa Manchón
(University of Murcia),
Paul Meara
(Emeritus Professor, University of Swansea),
Bernie Mohan
(Emeritus Professor, University of British Columbia, & Honorary Research Fellow, Kings College, London),
Rhonda Oliver
(Curtin University, Perth),
Lourdes Ortega
(Georgetown University),
Virginia Samuda
(Non-aligned),
Peter Skehan
(Birkbeck College, University of London),
Remi Van Compernolle
(Carnegie Mellon University),
Ineke Vedder
(University of Amsterdam).

Subject index

P

4/3/2 procedure 43–49, 56, 60

A

accuracy 6, 7, 8, 10, 20, 21, 33, 34, 45–49, 58, 66–7, 77–79, 97–99, 101, 102, 121, 144–146, 149, 150, 153–157, 159–162, 165, 170–179, 182, 185–189, 202, 204, 205, 206, 209, 212, 217, 280, 281, 290.

activity 226–228, 248, 255–261, 263, 269, 275

activity theory 167, 221, 227, 251, 256, 258–260, 263, 275, 276, 277, 308, 317

affordances 227–229, 247–249, 280, 285, 304, 305, 314, 315, 317, 324

appropriation 228, 234, 244, 245, 247

assessment 99, 100, 145

B

backstage talk 228, 238–240, 243, 247, 250

blocking 31, 32

C

CAF (complexity, accuracy, fluency) 9, 35, 48, 76, 78–81, 97, 98, 99, 102, 118, 143, 149, 174, 176, 178, 182, 183, 187, 223, 281, 322–324

co-construction 90, 228, 255, 256, 258, 267

collaborative theory 146, 147

comparator recording 175, 193, 196, 218

Complex Dynamic Systems Theory (CDST) 311–317, 319, 320–324

complexity 99, 105, 122, 126, 127, 129, 133–134, 136–139, 144, 145, 149, 152, 154, 155, 157–165, 171–176, 178, 179, 180, 182–184, 186–190, 204, 217, 280, 286,289, 290, 305

comprehension strategies 265, 268, 276

conference presentation task 203

content repetition 49, 60, 76–79, 173, 190, 317

D

declarative knowledge 27, 29, 30, 33–36, 38, 47, 172

deliberate practice 39

delivery 219, 235, 249

discourse competence 99–101

discourse performance 97, 99–102, 104, 106, 108, 109, 111, 112

distribution of practice 27, 28, 32, 37

E

effect of interlocutor 120, 141

emergence 284, 286, 304, 313, 315n2, 324, 328, 329

emic 311, 317

English for Academic Purposes (EAP) 193, 196, 206, 210, 220

English for Specific Purposes (ESP) 69, 193, 196, 203, 214, 222, 225

enhanced repetition 178, 193, 194, 196, 206, 217, 218

ethnographic (methods) 223, 224, 226, 229, 251

expertise 11, 12, 23, 39, 225

F

feedback 35, 93, 108, 110, 119, 120, 128, 138, 139, 175, 190, 193, 196, 197, 201, 203–206, 209, 216–219, 231, 233n1, 240, 245, 246, 248, 250, 258, 262, 276, 281, 283, 288, 314

fluency 28, 32, 33, 35, 36, 43–49, 56–60, 63, 64, 66–67, 75–81, 97–100, 102, 105, 118, 120, 143–145, 149, 150, 153, 155–162, 164–165, 171, 172, 174–176, 178, 179, 182, 185–189, 202, 204, 217, 223, 240, 279–281, 285, 287, 289–291, 296, 301, 306, 312, 322

formulaic speech 312

frontstage talk 228

G

goal-setting 279, 283, 289, 295, 297, 298, 302

H

high/low criterion 143, 147, 150–153, 158

I

implementation factors 80, 119

improvisation 249

inter-cultural task repetition 117, 118, 122, 138

interleaving 27, 31–33

intra-cultural task repetition 122, 138

iteration/task iteration 43, 47, 49, 53–59, 62–67, 99, 145, 284, 285, 304, 311, 312, 316–318, 323–325

L
language-related episodes (LRE) 117, 119–122, 125–140, 317
learning paths 311, 318
Levelt's speech processing model 9, 11, 46–47, 76, 101, 105, 106, 171–172, 280–281, 111, 280
lexical production 81, 93
limited attentional resources model 136, 143, 165,
listen-and-do task 255, 260
low/high criterion 143, 147, 150–153, 158

M
mean length of run (MLR) 153, 157, 158, 161, 164
monologic narrative task 171

N
non-appropriation 245
nonlinearity 311, 323, 324

O
other-regulation 255, 257, 258, 260, 275

P
paraphrasing 242
picture story 44, 46, 57, 60, 61, 64, 65, 67, 71, 179–181
planning 27, 38, 47, 58, 65, 68, 70, 80, 83, 93, 101–103, 105, 108, 110–111, 118, 146, 171–173, 176, 217–219, 224, 249, 251, 296
poster carousel 78, 120, 196, 200–203, 205, 206, 210, 214–217, 229–231
practice 27–33, 35, 37, 47, 82, 97, 145, 162, 165, 173, 175, 200, 201, 206, 213, 216, 219, 227, 235, 236, 242, 287, 306, 311–313, 315, 320
pre-casts 204
procedural knowledge 29, 30, 33–35, 37, 205
proceduralisation 206

R
recursive cycles 283
reflection activities 193
rehearsal 145, 204–206, 219, 228, 320
repeated practice 27, 31–33, 47, 312, 313
replication study 193
retrieval speed 313
revoicing 235, 247, 250
risk-taking 223

S
scaffolding 90, 100, 225, 229, 249, 255, 256, 258, 260, 267, 273, 276
scenario (task) 196, 206, 207, 209, 218
self-evaluation 279, 289–291, 295, 297, 298, 302
self-regulation 255, 257, 258, 260, 275, 276, 279, 282–284, 285–289, 303–309
semi-structured interviews 223
sociocultural theory 142, 177, 222, 251–256, 277, 309
specific task repetition 43, 46, 64, 67, 76
stimulated recall 34, 78, 140, 171, 176, 177–181, 183–192
student agency 248

T
task complexity 75, 76, 80–82, 85–89, 91–96,
task performance criteria 143, 146, 147, 150, 165
task recycling 79, 145, 146, 193
task rehearsal 145, 320
task type 44, 45, 56, 66, 67, 77, 78, 94, 99, 143, 144, 148, 150, 151, 154–156, 158, 159–162, 165, 167, 307
task-as-process 28, 36, 46, 66, 247, 284, 312, 316
task-as-workplan 28, 36, 37, 46, 66, 193, 214, 247, 255, 284, 285, 304, 316–319
teacher-student interaction 255
time pressure 41, 43, 45, 47–49, 56, 57, 64–68, 70, 99, 116, 176, 192
trade-off effect 143, 162, 173, 188
trajectory/trajectories 223, 226, 231, 233, 252, 288, 291, 319,
transfer/transfer of learning 27, 28, 33–36, 45, 151, 171, 173, 174, 179, 187, 188, 190, 311, 320, 321, 324
Transfer Appropriate Processing (TAP) 34, 35, 173,
transformation 311, 321, 324
trigrams 43, 45, 50, 51, 56, 59–64, 66, 67
types of task repetition 79, 89, 91, 102, 118, 173, 190
- exact (task repetition) 75–77, 79–81, 83, 85–89, 9193, 99, 139, 188, 231
- external 89, 202, 257, 282, 284, 288, 305, 314, 324
- immediate 38, 43–45, 47–49, 56, 60, 61, 98, 102–104, 110–112, 252, 308
- internal/within-task 189, 193, 196, 282, 283, 305, 306, 312
- procedural 75–79, 81–93, 120, 140, 173, 176, 179, 190, 226, 231, 244, 250, 281, 282, 285–287, 297, 317, 324

V
video-based narrative 103

After more than 20 years of research, this is the first book-length treatment of second language task repetition – the repetition of encounters with a task that involve re-using the same content with the same overall purpose. The topic links task performance with the growing mastery of both the task and of relevant language, and constitutes a site with special potential to promote learning within and across language lessons, and for preparing students for assessment and of course real-world language performance. The volume assembles chapters that complement each other in interesting ways: significant background reviews, studies of patterns of change across task repetition iterations, and reports on the use and nature of task repetition in language classes in on-going programmes. Contributors draw on a variety of interpretive frameworks and report from a range of language educational contexts. The volume will be of interest to language researchers, teacher educators, teachers, and students, as well as others interested in the contribution of task repetition to learning.

ISBN 978 90 272 0114 0

John Benjamins Publishing Company