Lang. Teach. (2019), **52.3**, 374–384 © Cambridge University Press 2017
doi:10.1017/S0261444817000283 First published online 4 October 2017

Replication Research

Replication in task-based language teaching research: Kim (2012) and Shintani (2012)

Andrea Révész University College London
a.revesz@ucl.ac.uk

This paper calls for the replication of two task-based language teaching (TBLT) studies, Kim (2012) and Shintani (2012). Kim contributes to the large body of TBLT research on output-based tasks, whereas Shintani initiates research into input-based tasks, a less well researched domain. These studies deserve to be replicated for several reasons. Unlike the bulk of existing TBLT research, they investigate task-based second language (L2) development in addition to performance, and are conducted in actual classroom rather than laboratory contexts. Also, they are well designed studies, including detailed descriptions of the instruments and methodological procedures, making replication feasible. First, this paper provides an overview of the studies. Next, approaches to approximate and conceptual replications are suggested in order to assess the internal and external validity of the original experiments.

1. Introduction

Task-based approaches to language teaching have received growing attention from both educators and researchers over the past few decades. In a variety of settings, the construct of task is increasingly adopted as either the principal (Long 1985; Van den Branden 2006) or one of the key units of syllabus and assessment (Bygate 2000; Ellis 2003). An often-cited definition describes tasks as activities 'where meaning is primary; there is some communicative problem to solve; some sort of relationship with real-world activities; and the assessment of task is in terms of a task outcome' (Skehan 1998: 95). There are both theoretical and practical rationales underlying the use of tasks in L2 teaching and assessment. First, tasks provide a platform for meaningful L2 use while simultaneously offering opportunities for a focus on form (Long 1991), a combination regarded as ideal for L2 developmental processes to unfold (Long & Robinson 1998). Second, the principles of TBLT correspond well to notions such as learning-by-doing and student-centred teaching, which are widely promoted in the field of general education (e.g., Dewey 1913/1975). Finally, TBLT advocates using a task-based needs analysis as a basis for syllabus design, to ensure that learners engage in language practice that reflects their real-life academic, professional, and/or personal goals (Long 2005). As a consequence, TBLT is likely to create a more motivating learning environment than more traditional approaches that often lack face validity with L2 learners. In light of

https://doi.org/10.1017/S0261444817000283 Published online by Cambridge University Press

the psycholinguistic and practical benefits that task-based approaches offer, much research has been carried out in recent years to explore ways to maximise the effectiveness of TBLT (for review, see Ellis 2003; Samuda & Bygate 2008; Long 2015).

With the ultimate goal of informing task-based lesson planning and syllabus design decisions, the most prolific line of TBLT research has involved investigating the effects of task manipulations on interactional patterns and linguistic output during task performance at one point in time. So far, few studies (e.g., Bygate 2001; Révész 2009; Kim 2012; Baralt 2013) have directly investigated the impact of manipulating task variables on longer-term gains in L2 knowledge and/or processing skills. Although cognitive-interactionist frameworks of task-based performance and development propose that task affects identified for output production (Skehan 1998; Robinson 2001) and interactional patterns (Robinson 2001) will transfer to L2 developmental outcomes in predictable ways, only research that directly establishes such relationships can offer firm evidence of this. As the aim of TBLT is to promote L2 development, it appears imperative to replicate and expand on the few studies that have begun to examine how task-related variables influence L2 learning. This paper will propose replication of one such study, Kim (2012), published in *Studies in Second Language Acquisition*. Among the existing studies on the effects of task manipulations on linguistic performance and development, this study appears particularly worthy of replication for several reasons. First, it was conducted in a real classroom rather than in a laboratory context, hence it has high ecological validity. Second, the task manipulations investigated were found to be successful in promoting interaction-driven language opportunities as well as L2 development. Thus, if the results are replicated, insights from the study could be used as a basis for providing guidance to teachers on how task design features can be altered to facilitate interaction and learning. Last but not least, the study has a solid and rigorous research design, which is described in sufficient detail to allow for replication.

Besides the paucity of developmental studies, another limitation of current TBLT research lies in its predominant focus on output-based tasks, that is, tasks that require learners to produce output through speaking or writing. To date, little research effort has been dedicated to input-based tasks, which engage learners in listening or reading without imposing the need on learners to generate output during task-based work (Shintani 2012). Lack of research on input-based tasks represents a key gap in TBLT research: in the TBLT framework, input-based tasks constitute the primary source of comprehensible input (Shintani 2016), which is generally considered to be a prerequisite for L2 development to occur. Shintani's (2012) study, published in *Language Teaching Research*, is one of the first to explore the potential of input-based tasks to foster L2 development. Approximate and conceptual replications of this research would appear a good way to contribute to this under-researched domain. Like Kim (2012), the study boasts high ecological validity as it was carried out in an actual classroom setting; its findings are promising and thus provide a potentially good basis for drawing pedagogical implications; and the research is again well designed and clearly described, which facilitates replication. An additional positive feature of the study is that the participants are young learners, a population that has received relatively little attention from L2 researchers.

The next section of this paper provides a description of the background, methodology and findings of the two studies. Then suggestions are made as to how to approach their replication. The last section offers a brief conclusion.

https://doi.org/10.1017/S0261444817000283 Published online by Cambridge University Press

2. The original studies and suggested approach to replication

2.1 Kim (2012)

The aim of Kim's (2012) study, situated in the cognitive-interactionist paradigm, was to examine how task complexity may influence the incidence of interaction-driven language learning opportunities and L2 development in a classroom context. The role of task complexity, defined as the inherent cognitive demands posed by a task, has been the subject of a large amount of research in the field of TBLT. This line of research has been inspired by two cognitive-interactionist models for TBLT and learning: Robinson's (2001) Cognition Hypothesis and Skehan's (1998) Limited Capacity Model. The primary concern of both frameworks is to explain how task characteristics can affect L2 developmental processes and outcomes, with the practical goal of informing teachers' decisions about task grading and sequencing when they plan task-based lessons and syllabuses. Thus far, the bulk of task complexity research has been concerned with investigating the effects of task complexity on the linguistic performance areas of complexity, accuracy and fluency. Kim's research was among the first studies (Révész 2009; Baralt 2013) to examine the effects of task complexity on L2 development by testing one claim of the Cognition Hypothesis, namely, increasing the cognitive and conceptual demands of tasks leads to greater amount of interaction and higher occurrence of feedback and thereby promotes L2 development. Kim's research is unique in that it was conducted in a classroom setting, where the tasks were part of the syllabus students followed rather than provided by the researcher for the purposes of the study.

The study employed a pretest-posttest-delayed-posttest design, spanning six weeks. The participants took the TOEIC Bridge test in Week 1, the pretest was administered in Week 2, the treatment took place in Weeks 3–4, and the posttest and delayed posttest were given in Weeks 5 and 6 respectively. The participants were 191 adult Korean learners, who majored in various subjects but all studied English as a foreign language at a university in South Korea. The mean age of the participants was 18.82 (SD = 1.08), and their proficiency levels ranged from beginner to intermediate as determined by their TOEIC Bridge scores.

Four intact classes were selected for the experiment and the participants were randomly assigned to three experimental groups and a comparison group. The overall English proficiency of the groups was found to be comparable. During a semester, the experimental groups received task-based instruction, and carried out 12 tasks. Of these, the study included four convergent tasks in the treatment, all focusing on the theme of work, reflecting the students' needs and interests. The four tasks had the following topics: finding a part-time job, working as a matchmaker, discussing promotion opportunities, and hiring employees. The tasks varied in complexity along the +/−reasoning and +/−few elements dimensions (Robinson 2001) across the three experimental groups. The simple group completed information-transfer tasks (−reasoning). The +complex and ++complex groups both carried out decision-making tasks (+reasoning), but differed as to the number of elements participants needed to consider to make a decision. The ++complex group had to take into account more factors (−few elements) than the +complex group (+few elements). During task performance, the learners worked in self-selected pairs, with the same partner in all four tasks. The

https://doi.org/10.1017/S0261444817000283 Published online by Cambridge University Press

comparison group received traditional instruction organised around the same topic, but had fewer opportunities to engage in learner–learner interaction.

Pretest to posttest development was determined in terms of participants' advancement along the six developmental stages established for question formation by Pienemann & Johnston (1997). Three tests were used to elicit questions: two individual and one paired oral production test. Interaction-driven language learning opportunities were operationalised as language-related episodes (LREs) (Swain & Lapkin 1998) involving questions. Once the LREs had been identified in the transcripts of participants' treatment task performance, the total number of LREs and the number of LREs involving developmentally advanced questions (Stages 4 and 5) were calculated for all four tasks combined.

The results were found to be largely in line with the predictions of the Cognition Hypothesis, with increased task complexity having a positive impact on both the incidence of language learning opportunities and L2 development. As regards language learning opportunities, a one-way ANOVA revealed that the ++complex group generated the greatest number of LREs involving questions, and the +complex group produced more LREs as compared to the simple group. The occurrence of LREs involving developmentally advanced questions was also the highest for the ++complex group, but the ++complex group only produced significantly more questions than the simple group. Turning to the developmental results, a chi-square test found a significant link between task complexity and improvement in question formation, operationalised as increase versus no increase in developmental stage from the pretest to the posttests. Follow-up analysis of the adjusted standardised residual revealed that the ++complex group outperformed the comparison group but there was no significant difference in development among the +complex, simple, and comparison groups.

2.1.1 Approach to replication

Given the overall positive results of Kim (2012), a series of replications would be helpful to ascertain the generalisability of the findings of the study. A first approach to replication research might usefully take the form of keeping the methodology the same but conducting a more detailed analysis of the new dataset, to shed light on whether the results can be extended to more specific interactional features, other linguistic constructions, and general measures of linguistic performance. As Kim herself acknowledges, it would be interesting to adopt a more detailed approach to analysing the interaction-driven language learning opportunities that arose in learner–learner interactions during the treatment. In addition to identifying LREs, the transcripts could be coded for specific feedback types such as recasts, confirmation checks, comprehension checks and clarification checks. This would increase the comparability of the findings to other studies investigating the effects of task complexity on the occurrence of interactional features (e.g., Robinson 2007; Gilabert, Barón & Llanes 2009; Révész 2011), and would provide additional insights about the relationships between task complexity, the incidence of specific feedback types during peer interaction, and L2 development. Another useful way to explore a dataset collected through a replication would be to extend the analysis to other linguistic constructions. As suggested by Kim, coding for physically non-salient and communicatively redundant features would be particularly valuable, since non-target-like

https://doi.org/10.1017/S0261444817000283 Published online by Cambridge University Press

utterances involving such constructions are less likely to generate communication breakdowns and, therefore, might be less affected by differences in task complexity. Researchers should also consider analysing the data from replication in terms of the general linguistic performance measures of complexity, accuracy and fluency (CAF) typically employed in TBLT research. This would not only enable comparing the results of the study with the large body of research that exists on the effects of task features on the quality of speech production but would also help extend this line of research to investigating developmental changes in CAF indices as a function of task complexity.

A second useful avenue for approximate replication research would be to address the same research questions but alter the design slightly to further elucidate the results of the original study. One methodological addition that would assist in increasing the internal validity of the study would be to include an independent measure of task complexity. This would help ensure, as Kim also points out, that the tasks versions which were designed to be more complex did indeed exert greater cognitive demands on the learners (Norris 2010; Révész 2014). Among the various methods available to assess task-generated cognitive demands (Révész 2014), learner perception questionnaires and expert judgements would appear good options in classroom research, as there is growing evidence that these methods offer valid measurements while being non-intrusive and easy to administer (Révész, Michel & Gilabert 2016; Sasayama 2016). A second minor methodological change might involve manipulating the treatment tasks along a single rather than multiple task complexity dimensions. In Kim's research, task complexity was increased along two variables simultaneously, +/−reasoning demands and +/−few elements. In a replication study, it would be helpful to disentangle the effects of these two task factors by operationalising different levels of task complexity in terms of only one of these dimensions at a time. A third way to modify the study to enhance its construct validity might be to increase the number of participating classes and assign more than one intact class to the three experimental and comparison conditions. In this way, any class or teacher effects could be avoided. Alternatively, if practical constraints permit, a future approximate replication study could employ a split class design, in which pairs of students in each class are randomly assigned to one of the four conditions (Marsden 2007). This option would appear easier to implement if the comparison group engaged in an equivalent amount of oral (but non-task-based) interaction to that of the experimental groups. This would independently be desirable to make the groups more comparable.

A third fruitful area for approximate replication research would be to keep the design of the original study, but change the participating learner populations, educational context, task types, or modality, in an attempt to assess whether the results obtained by Kim (2012) would be generalisable across these factors. Kim's study was conducted at a university in South Korea where participants were young adults who studied English as a foreign language (EFL). It is important to examine whether the findings can be replicated with learners of different ages and learners studying in other institutional settings, including primary and secondary schools. These populations, for example, might have differential motivational characteristics and levels of willingness to communicate in an L2, which could affect the link between task complexity and the interactional patterns in which they engage (Robinson 2001). Approximate replications in English as a second language contexts are also warranted,

https://doi.org/10.1017/S0261444817000283 Published online by Cambridge University Press

as this would permit comparisons with groups of learners who do not share the same first language (L1) and therefore might be more likely to experience communication breakdowns during task-based work (Gass & Varonis 1985). In addition, it is important to consider the extent to which the findings are generalisable to various task types such as narrative and description tasks, given that they would likely elicit discourse features different from the ones that emerge in information-transfer and decision-making tasks. A related area for approximate replication is to see whether manipulations along other task dimensions in Robinson's and Skehan's frameworks (e.g., planning time) would yield similar results.

Besides approximate replications, conceptual replications employing verbal protocol methodology would also be valuable to gain insights into the cognitive processes in which participants engaged while carrying out the treatment tasks. Among the introspective methods, the addition of stimulated recall methodology (Gass & Mackey 2016) would lend itself particularly well to this study, since the stimulated recall technique, although subject to memory decay, does not interfere with speech production processes. The addition of verbal protocol data would enable researchers to assess the validity of some of the explanations offered by Kim in the original study regarding the effects of task complexity on attentional processes and depth of interlanguage analysis.

2.2 Shintani (2012)

Shintani's (2012) study examined the extent to which engaging in input-based tasks can promote interaction and L2 development among young L2 learners in a TBLT context. In doing so, the study helped address an oft-cited criticism of TBLT: given its exclusive focus on production, it provides learners with no opportunities to get exposure to new language, which is a necessary condition for L2 development to occur (Swan 2005). Indeed, existing research on TBLT has predominantly focused on production-based tasks (see, however, Ellis et al. 1999), even though TBLT courses for beginning language learners inevitably need to be organised primarily around input-based tasks, as learners do not yet have the L2 knowledge and skills to produce the L2 in a meaningful way (Ellis 2009). In response to this research gap, Shintani (2012) launched an investigation into the potential of one type of input-based task, listen-and-do activities, in facilitating L2 learning opportunities and acquisition. In particular, it explored (1) the nature of the teacher–learner interactions that arise during task-based work, and (2) the extent of lexical and grammatical development that results from engaging in listen-and-do tasks.

The participants were 30 six-year-old children, all Japanese beginner learners of EFL. They were studying in a private language school in Japan. The study used a quasi-experimental design, including a pretest and two posttests. The children were assigned to an input-based task group and a control group, with the input-based task group being further divided into a class of six and nine learners. The input-based task group participated in eight task-based lessons over five weeks and performed three input-based tasks in each lesson. The control group was included to establish any test-practice effects and only took part in the testing sessions. It learned English songs, received Total Physical Response Instruction and practised the use of the English alphabet. Both groups were taught by the researcher, an experienced

https://doi.org/10.1017/S0261444817000283 Published online by Cambridge University Press

EFL teacher. The pretest was administered two weeks before the instruction began, and posttests 1 and 2 were conducted one and five weeks after the last lesson respectively. Each test involved the completion of a series of assessments, comprising comprehension as well as production-based tests.

The three listen-and-do tasks that participants performed in each of the eight lessons were designed to facilitate the acquisition of 36 vocabulary items (24 nouns and 12 adjectives) and the plural *-s* construction (presented with 6 nouns). The plural *-s* was targeted indirectly and in a reactive manner through the provision of feedback in response to comprehension errors and through answers to learners' questions about the meaning of vocabulary items. The first task was contextualised as a game, and involved the learners in listening to the teacher's commands and selecting picture cards denoting the target linguistic features. The children were told that, in order to help the zoo or the supermarket, they needed to find the appropriate cards and put them in their place. In the second task, each participant held the same picture cards as in Task 1, and was asked to identify cards that matched the teacher's statements. The picture cards represented animals and objects, and the non-linguistic aim of the task was to help the animals find the objects. The third task was a Picture Bingo, using the same picture cards as Tasks 1 and 2. The lessons were conducted primarily in English, but learners were allowed to use Japanese if they felt it necessary. All the lessons were both audio- and video-recorded.

Development in vocabulary knowledge was assessed by two comprehension- and two production-based tests. One of the comprehension and production tests was task-based, whereas the other contained discrete items for both test types. Changes in the children's ability to comprehend the plural construction was gauged by a multiple-choice comprehension test, and their ability to produce the plural *-s* form was measured by the Wug test (Berko 1958) and a Same or Different task. During the latter the researcher asked questions from the children to elicit the use of the target form. The authors carried out a series of parametric and non-parametric within- and between-subjects statistical tests to assess the significance of any changes in children's knowledge of vocabulary and grammar due to the treatment.

Analysis of learners' self-directed and other-directed speech during the listen-and-do tasks revealed that both types of discourse fulfilled roles that are considered facilitative of L2 learning. Self-directed speech served as a means of self-regulation and language play, whereas other-directed speech led to the occurrence of negotiation of meaning and focus-on-form episodes. For the acquisition of vocabulary, the statistical analyses yielded an advantage for the input-based task group over the control group on all assessments for both target nouns and adjectives. The experimental group also outperformed the control group on both posttests, despite the fact the control group achieved a higher score on one of the comprehension tests at the time of the pretest. Finally, participants in the input-based task group showed superior pretest-posttest gains in the use of the plural *-s* on the comprehension test. They also achieved higher scores than the control group on both posttests, while having comparable results at the grammar pretest. No significant differences were found between the input-based task and control groups on the grammar production tasks. It is worth noting, however, that two learners achieved gains in productive knowledge of the plural *-s* in the experimental group, whereas none of the control group participants showed evidence of improvement in the productive use of the target grammatical construction.

https://doi.org/10.1017/S0261444817000283 Published online by Cambridge University Press

2.2.1 Approach to replication

In light of the scarcity of research on input-based tasks and the careful design and encouraging results of Shintani (2012), it would appear useful to replicate the study to substantiate the findings and assess their generalisability. As a start, a series of approximate replication studies are proposed to address the methodological limitations of the original research. Shintani herself notes that the relatively small sample size puts a limit to the generalisability of the findings. Replicating the experiment with a considerably larger number of students from the same population would allow for more robust statistical analyses and thereby help ascertain the validity of the research. Another limitation of Shintani (2012) has to do with the fact that the researcher also served as the teacher of the participating classes. It would be worthwhile to repeat the study with these two roles separated, the researcher acting as an observer rather than a participant, thereby reducing the chance of inadvertent researcher bias towards the results. To further increase construct validity, ideally more than one teacher would be recruited, each teacher leading at least two classes, one assigned to the experimental treatment and one to the control condition, to lessen a possible teacher effect. A replication would also benefit from incorporating a comparison group, who is exposed to input-based but non-task-based teaching. In this way, it would be possible to isolate the impact of using tasks in input-based instruction. Finally, the inclusion of a greater number of plural items in the input (i.e., using more flashcards depicting multiple items) would provide learners with more opportunities to develop knowledge of this construction.

An additional useful set of approximate replication studies, as Shintani also notes, would focus on exploring whether the findings obtained transfer to other task types, populations, contexts, and linguistic targets. Shintani exclusively investigated listen-and-do tasks; thus, it is unclear the extent to which other input-based tasks would succeed in promoting L2 development. It would be particularly helpful to examine the learning potential of different comprehension-based input tasks which also require a response but do not ask learners to perform physical actions to demonstrate that they have processed the input. Approximate replication studies are also warranted with different learner populations. Shintani's study included young Japanese learners in a private kindergarten context, with the parents being invested in the children's education. It is important to investigate whether the treatment would equally benefit children from different L1 backgrounds, less educationally motivated families, and state kindergarten settings where groups tend to be larger. Similarly, studies with older learners are also needed; replications involving older populations would require modifications to the listen-and-do tasks to make sure that the topics of the tasks are age-appropriate and, hence, of interest to the students. Replicating the study with different grammatical constructions would also assist in establishing the effectiveness of input-based tasks in L2 learning. In their current form, the listen-and-do tasks in Shintani's study only lend themselves to the teaching of meaning-bearing items with some communicative value. Meaning-bearing items, however, might differ in physical salience. Thus, it would be worth exploring how the physical salience of the target construction may influence the success of the treatment, keeping communicative value constant. It may well be that linguistic constructions that are more physically salient than the plural *-s* would be noticed by learners with greater ease, leading to more substantial L2 gains.

https://doi.org/10.1017/S0261444817000283 Published online by Cambridge University Press

A third group of replication studies, involving more than one or two minor changes to the original design, are also necessary to learn more about the capacity of input-based tasks to foster development in L2 lexis and grammar. Although input-based are particularly important in the initial stages of task-based instruction (Ellis 2009), they are also key sources of new linguistic material in more advanced TBLT courses. Thus, it would appear beneficial to replicate Shintani (2012) with learners from higher levels of proficiency. This would require not only altering the participating student population but also the target linguistic features and the tasks that learners perform. More advanced language learners need to acquire more abstract lexical and grammatical constructions than the plural *-s*, which are unlikely to be teachable through listen-and-do tasks relying on pictorial input. Comprehension-based and enriched-input tasks would seem more suitable for teaching and learning abstract features by offering a more convenient format for using textual input (oral or written). In addition, as with Kim (2012), there would be value in examining whether the results of Shintani (2012) could be replicated in an online context. Transferring the study to an online environment would involve several modifications to the original research, such as changing the format of the feedback provided, reconsidering the role of the teacher, and recruiting young learners with sufficient computer literacy to carry out the tasks.

Like for Kim (2012), conceptual replications that attempt to gain insights into the learners' perspectives about the treatment would be helpful. Given the oral modality of the input, adding stimulated recall interviews would appear an appropriate way to tap the children's thought processes during task-based work. The incorporation of individual and focus groups interviews could help gain additional insights into children's perceptions about the usefulness of engaging in listen-and-do tasks and the extent to which they enjoyed performing them.

3. Conclusion

The role of tasks in L2 learning and teaching has received much attention from researchers in recent years. So far, however, the bulk of research has focused on task-based performance; only a few studies exist that have looked into how L2 development may be facilitated in task-based contexts. Also, relatively few task-based studies have been conducted in actual classroom settings, even though the aim of TBLT research is to inform L2 pedagogy. The two studies recommended for replication here are both among the initial attempts to address these gaps in the literature. Kim (2012) made a significant contribution to the large body of research examining output-based tasks (Plonsky & Kim 2016), whereas Shintani (2012) moved the field forward by looking into the language learning potential of input-based tasks, an area which is less explored to date. These studies are good candidates for replication not only because of their theoretical and pedagogical significance but also because they are well designed, the data collection and analytical procedures are clearly described, and the research instruments employed are easily accessible (visit www.iris-database.org/iris/app/home/index). This paper has suggested several ways in which approximate and conceptual replications of the original studies could potentially provide further evidence for the validity of the studies, assess the generalisability of their findings, and raise our understanding of learners' internal processes during task-based work.

https://doi.org/10.1017/S0261444817000283 Published online by Cambridge University Press

References

Baralt, M. (2013). The impact of cognitive complexity on feedback efficacy during online versus face-to-face interactive tasks. *Studies in Second Language Acquisition* 35.4, 689–725.

Berko, J. (1958). The child's learning of English morphology. *Word* 14.2–3, 150–177.

Bygate, M. (2000). Introduction to special issue: Tasks in language pedagogy. *Language Teaching Research* 4.3, 185–192.

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*. Harlow: Longman, 23–48.

Dewey, J. (1975). *Interest and effort in education*. Carbondale: Southern Illinois University Press, Arcturus Books. (Original work published 1913)

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

Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics* 19.3, 221–246.

Ellis, R., R. Heimbach, Y. Tanaka & A. Yamazaki (1999). Modified input and the acquisition of word meanings by children and adults. In R. Ellis (ed.), *Learning a second language through interaction*. Amsterdam/Philadelphia: John Benjamins, 23–48.

Gass, S. M. & E. Varonis (1985). Task variation and non-native/non-native negotiation of meaning. In S. Gass & C. Madden (eds.), *Input in second language acquisition*. Rowley, MA: Newbury House, 149–161.

Gass, S. M. & A. Mackey (2016). *Stimulated recall methodology in applied linguistics and L2 research*. New York: Routledge.

Gilabert, R., J. Barón & A. Llanes (2009). Manipulating cognitive complexity across task types and its impact on learners' interaction during oral performance. *International Review of Applied Linguistics in Language Teaching* 47.3–4, 367–395.

Kim, Y. (2012). Task complexity, learning opportunities and Korean EFL learners' question development. *Studies in Second Language Acquisition* 34.4, 627–658.

Long, M. H. (1985). A role for instruction in second language acquisition: task–based language teaching. In K. Hyltenstam & M. Pienemann (eds.), *Modelling and assessing second language development*. Clevedon: Multilingual Matters, 149–161.

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, D. Coste, R. Ginsberg & C. Kramsch (eds.), *Foreign language research in cross-cultural perspectives*. Amsterdam, The Netherlands: John Benjamins, 39–52.

Long, M. H. (2005) (ed.). *Second language needs analysis*. Cambridge: Cambridge University Press.

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

Long, M. H. & P. Robinson (1998). Focus on form: Theory, research, and practice. In C. Doughty & J. Williams (eds.), *Focus on form in classroom second language acquisition*. Cambridge: Cambridge University Press, 15–41.

Marsden, E. (2007) Can experiments both inform theory and practice? *British Educational Research Journal* 33.4, 565–588.

Norris, J. M. (2010). *Understanding instructed SLA: Constructs, contexts, and consequences*. Plenary address delivered at the annual conference of the European Second Language Association (EUROSLA), Reggio Emilia, Italy, September.

Pienemann, M. & M. Johnston (1987). Factors influencing the development of language proficiency. In D. Nunan (ed.), *Applying second language acquisition research*. Adelaide, Australia: National Curriculum Resource Centre, 45–141.

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

Révész, A. (2009). Task complexity, focus on form, and second language development. *Studies in Second Language Acquisition* 31.3, 437–470.

Révész, A. (2011). Task complexity, focus on L2 constructions, and individual differences: A classroom-based study. *The Modern Language Journal* 95.S1, 162–181.

Révész, A. (2014). Towards a fuller assessment of cognitive models of task–based learning: Investigating task-generated cognitive demands and processes. *Applied Linguistics* 35.1, 87–92.

https://doi.org/10.1017/S0261444817000283 Published online by Cambridge University Press

Révész, A., M. Michel & R. Gilabert (2016). Measuring cognitive task demands using dual task methodology, subjective self-ratings, and expert judgments: A validation study. *Studies in Second Language Acquisition* 38.4, 703–737.
Robinson, P. (2001). Task complexity, cognitive resources, and syllabus design: A triadic framework for investigating task influences on SLA. In P. Robinson (ed.), *Cognition and second language instruction*. New York, NY: Cambridge University Press, 287–318.
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* 45.3, 193–213.
Samuda, V. & M. Bygate (2008). *Tasks in second language learning*. Basingstoke: Palgrave Macmillan.
Sasayama, S. (2016). Is a 'complex' task really complex? Validating the assumption of cognitive task complexity. *The Modern Language Journal* 100.1, 231–254.
Shintani, N. (2012). Input-based tasks and the acquisition of vocabulary and grammar: A process–product study. *Language Teaching Research* 16.2, 253–279.
Shintani, N. (2016). *Input-based tasks in foreign language instruction for young learners*. Amsterdam, The Netherlands: John Benjamins.
Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.
Swain, M. & S. Lapkin (1998). Interaction and second language learning: Two adolescent French immersion students working together. *The Modern Language Journal* 82.3, 320–337.
Swan, M. (2005). Legislation by hypothesis: The case of task-based instruction. *Applied Linguistics* 26.3, 376–401.
Van den Branden, K. (ed.) (2006). *Task-based language teaching: From theory to practice*. Cambridge: Cambridge University Press.

ANDREA RÉVÉSZ is Senior Lecturer in Applied Linguistics at the UCL Institute of Education, University College London. Her research interests include the roles of tasks, input, interaction, and individual differences in instructed second language acquisition, as well as the cognitive processes underlying L2 performance. She is associate editor of the journal *Studies in Second Language Acquisition*, and serves as Vice-President of the International Association for Task-based Language Teaching.

https://doi.org/10.1017/S0261444817000283 Published online by Cambridge University Press