Abstract
This chapter provides an overview of key methods that are used to examine the role of tasks in second language performance and development. For each method, I provide a short description of the area(s) of research in which it is typically used, followed by examples to demonstrate how the method can be employed to investigate task-related issues. I also highlight and discuss the advantages and limitations associated with each method, and consider how potential limitations might be mitigated through careful design and implementation. Next, I turn to a discussion of some current issues in TBLT research methodology, such as the tension between internal and ecological validity, the need for more developmental and longitudinal research to complement the current focus on task-based performance, the value of investigating task-based processes besides products, the advantages and challenges of triangulating data sources, and the importance of thorough data reporting and transparency.

Methodological Approaches to Investigating TBLT: Advances and Challenges

Andrea Révész, University College London

In the area of instructed second language acquisition (SLA) research, the past three decades have seen a surge of interest in investigating tasks as a means of facilitating second language (L2) development. This increased interest has been motivated by a growing consensus among instructed SLA researchers that tasks have the capacity to create ideal circumstances for L2 learning by promoting the cognitive as well as social processes assumed to foster L2 development in instructed language learning contexts. Research attention to tasks has additionally been driven by the increasing acceptance of task-based language teaching (TBLT) as a valuable and feasible pedagogical approach to teaching second languages. As a result of the rising theoretical and practical importance of task-related research, a wide range of methods have been utilised by researchers, from laboratory experimental designs to classroom action research projects, to explore task-based learning. In this chapter, I will provide an overview of key methods that are used to examine the role of tasks in second language development. I will also highlight innovative approaches and methodological challenges in investigating task-based performance and learning.

1. Types of TBLT Research

1.1 Experimental and Quasi-experimental Research on Tasks and Task-based Programs

Much of the existing research on task-based performance and development has been quasi-experimental or experimental in nature. The primary focus has been to explore how the manipulation of task factors may influence the incidence of interaction-driven L2 learning opportunities, linguistic performance, and L2 development. More recently, researchers have also begun to examine how task-related variables may affect the cognitive processes that underlie task-based performance and learning (e.g., Kim, Payant, & Pearson, 2015; Révész, Kourtali, & Mazgutova, 2017; Torres, 2018). In this line of research, the independent variable is usually a task-related factor, whereas the typical dependent variables are linguistic outcome measures, interactional features, or process-oriented indices. Task-related factors that have received extensive attention are task types (e.g., narrative vs. decision-making, integrated vs.

independent), interactive task conditions (e.g., whether participants need or do not need to reach a consensus, see Ellis, 2003, and Mackey, 2012, for a review of interactive conditions), and task complexity (i.e., the inherent cognitive demands of tasks).

A study by Michel, Révész, Lu, Kourtali and Borges (in press) is a recent example of a study investigating task type effects. The researchers operationalised task type as the distinction between independent and integrated writing tasks. The independent task involved writing an essay, whereas, in the integrated task, participants were asked to produce a written summary of a listening and a written passage while synthesising the information from the two sources. Each participant completed two independent and two integrated tasks, the order of which was counterbalanced across participants. The dependent variables were the behaviours and associated cognitive processes of L2 writers, as captured by a variety of keystroke-logging and eye-tracking indices and qualitative comments gathered through stimulated recall protocols. The stimulated recall comments were elicited based on the last writing task participants had performed. As compared to the majority of previous studies on task type effects, a strength of this design was that the researchers included two rather than one version of each task type, which allowed for isolating the impact of task type from potential topic or prompt effects.

Lambert and Engler's (2007) research well illustrates how an experimental approach can be used to examine the impact of interactive conditions on L2 performance. The researchers utilised a 2x3 repeated-measures design, with goal orientation and information distribution as independent, within-subjects factors. The two levels of goal orientation were whether the task was open (i.e., the task did not have a predetermined outcome) or closed (i.e., the task had a pre-determined outcome). Information distribution was operationalised as having three levels: shared, one-way (i.e., one person holds all the information), and two-way (i.e., the information is split between participants). The order of the six conditions was counterbalanced, with each participant being exposed to all six conditions. The dependent variables were measures of linguistic complexity, accuracy, and fluency. A noteworthy feature of the design was that the researchers were able to generalise about the effects of various interactive conditions, given that these were investigated across three task types (ordering pictures, deciding responsibility, and arranging times).

While task type and interactive conditions have attracted considerable attention from TBLT scholars, most task-based research in the experimental paradigm so far has focused on the effects of task complexity on L2 performance. This area of research has largely been inspired by two cognitive-interactionist models of task-based learning, Robinson's (2001) Cognition Hypothesis and Skehan's (1998, 2009) Limited Capacity Model. These models make partially different predictions about how manipulations along certain task complexity dimensions will affect linguistic performance and development. With a view to testing these models, experimental studies of cognitive task complexity typically entail the following steps. First, researchers select or design a pedagogic task. Then, they develop two or more versions of the task with the intention that the versions differ in terms of cognitive demands along a particular task feature. For example, researchers might design two task versions, one posing more and the other imposing fewer reasoning demands. Next, researchers usually determine whether the task conditions have resulted in superior outcomes. To date, studies have primarily captured outcomes employing linguistic performance indices of complexity, accuracy, and fluency (CAF, Housen & Kuiken, 2009; Michel, 2017). Increasingly, however, TBLT scholars are also

concerned with investigating how cognitive complexity manipulations may affect L2 development in specific linguistic features, e.g., use of conditionals (Baralt, 2013; Kim, 2012; Kourtali & Révész, 2020; Nuevo, 2006; Révész, 2009; Révész, Sachs, & Hama, 2014; Torres, 2018), the frequency of language learning opportunities arising during interaction, e.g., negotiation of meaning and various types of feedback (Gilabert, Barón & Llanes, 2009; Kim, 2009; Révész, 2011), and the cognitive processes in which learners engage during task performance (e.g., Kim et al., 2015; Révész et al., 2017; Torres, 2018).

Until recently, one methodological weakness of task complexity studies has been that researchers assumed rather than substantiated the validity of their task manipulations (Norris & Ortega, 2003; Révész, 2014). In other words, they failed to provide independent evidence that the task version they constructed to be more complex did indeed exert greater cognitive demands on the learners. To deal with this shortcoming, a growing number of studies incorporate independent measures of cognitive complexity to ensure that the task complexity conditions reflect the intended experimental manipulation (e.g., Baralt, 2013; Malicka & Levkina, 2012; Révész et al., 2014; Zalbidea, 2017). To date, researchers have relied on a number of techniques to assess task-generated cognitive demands, including subjective self-ratings, subjective time estimations, dual-task methodology, eye-tracking, and expert judgments. Some scholars have even investigated and compared the usefulness of various methods to measure task-induced cognitive demands (e.g., Lee, 2019; Révész, Michel, Gilabert, 2016; Sasayama, 2016) with the aim of guiding validation work in future task complexity research.

Another key development in task-related experimental research has been a more sophisticated measurement of linguistic complexity, accuracy, and fluency, constructs which are often included as the primary dependant variables in task-based studies. For example, in response to calls to capture the dynamic and multidimensional nature of syntactic complexity (Bulté & Housen, 2012; Housen & Kuiken, 2009; Norris & Ortega, 2009), recent TBLT studies often include measures of phrasal, clausal and overall complexity rather than a single index of syntactic complexity. Similarly, TBLT researchers increasingly employ a variety of lexical diversity indices following recommendations in the literature (e.g., Jarvis, 2013).

While there have been many methodological advances in experimental task-related research, a gap that needs addressing includes a lack of studies that assess whether the findings obtained in experimental settings can be transferred to real classrooms. Although the experimental approach might lend itself best to laboratory studies, where researchers can control for a large array of potential confounding factors, it is also important to extend experimental research to real classroom settings. Otherwise, whether the findings obtained possess ecological validity remains unassessed. A few studies have examined the effects of task-related variables in actual classroom contexts. Kim (2012), for example, investigated how task complexity may affect L2 development and the incidence of interaction-driven language learning opportunities in Korean L2 English classrooms. Kim's research is noteworthy in that the tasks in which participants engaged came from the syllabus the students normally followed rather than being supplied and designed by the researcher for the purpose of the experiment.

Finally, it is worth highlighting, that the experimental research, albeit primarily focusing on task effects so far, can also be utilised to compare the TBLT approach as a whole with other types of

instructional options. Such comparative method studies are highly challenging to conduct, thus often suffer from methodological shortcomings, such as lack of pretesting, failure to include a control and/or a comparison group, absence of control over possible teacher and learner effects, the use of biased instruments towards one instructional treatment, and lack of evidence that the instruction was aligned with the intended methodological approach (Ellis, Skehan, Li, Shintani & Lambert, 2020; Ellis & Shintani, 2014; Long, 2015). Some recent comparative studies, however, have succeeded in avoiding many of these pitfalls. For example, De la Fuente (2006) investigated the relative effectiveness of the presentation-practice-production (PPP) approach and TBLT with or without explicit instruction. The study focused on the learning of L2 Spanish vocabulary items. Similar, Shintani (2013, 2015) compared the extent to which TBLT and PPP facilitated development in target vocabulary by Japanese child learners of L2 English. Shintani (2015) also examined the incidental learning of two grammatical features (plural –s, copula *be*). Besides having relatively robust designs, a strength of these three studies was the inclusion of process data (e.g., examination of interactional patterns), in addition to product-oriented pretest-posttest and pretest-delayed posttest measures (de la Fuente, 2006; Shintani, 2013, 2015).

1.2 Correlational/Associative Research on Learners and Tasks
Correlational, or associative, designs are another type of research that task-based scholars employ. Unlike experimental approaches, correlational designs do not involve manipulating variables with a view to establishing cause-effect relationships, but instead investigate associations among variables that remain unmanipulated. In task-based research, correlational designs have most frequently been used to explore how individual difference factors may relate to task-based outcomes. Typically, participants are measured in terms of an individual difference factor (e.g., anxiety, aptitude, creativity, motivation, working memory) and indices of linguistic performance, L2 learning, or interactional features assumed to drive L2 learning. In the next step, statistical procedures are used to identify associations between the two sets of variables.

A study by Dörnyei and Kormos (2000) was among the first correlational studies in the field of TBLT. The researchers set out to determine the relationship between task engagement and a group of motivational variables, social factors (e.g., group cohesiveness), as well as willingness to communicate in the participants' first language. Task engagement was operationalised as the number of turns and amount of speech produced by the learners. Self-report questionnaires were administered to measure participants with regard to the individual difference variables. To answer the research questions, the researchers computed correlations between the individual difference indices and the measures of task engagement. Several TBLT studies have adopted similar designs when examining how these and other individual difference factors, including working memory (e.g., Mackey, Adams, Stafford, & Winke, 2010) and creativity (Albert & Kormos, 2004; McDonough, Crawford, & Mackey, 2015), may influence task performance.

Increasingly, researchers are also using complex statistical techniques (e.g., structural equation modelling) to explore the role of individual differences in the context of TBLT. For example, a recent study by Wang (2019) aimed to identify the underlying facets of task motivation and task anxiety and how these factors relate to L2 motivation and foreign language and trait anxiety respectively. The study additionally examined the extent to which these motivational and anxiety-related factors predicted linguistic performance, expressed in terms of linguistic complexity, accuracy, and fluency indices. The researcher used motivation and anxiety

questionnaires to gain information about participants' motivational and anxiety profiles, and elicited linguistic performance data by means of a video narration task. Participants' responses to the questionnaires were first submitted to exploratory factor analyses. Next, structural equation modelling was conducted to examine the associations between task motivation and L2 motivation and between task anxiety, trait anxiety and foreign language anxiety. To address the links between the individual difference factors and the linguistic performance measures, a series of multiple regression analyses were carried out.

1.3 Aptitude-Treatment-Interaction (ATI) Research

The past decade has also seen a growing number of TBLT studies adopting the ATI research paradigm. The aim of ATI studies is to determine how individual differences among learners may moderate the effectiveness of various types of L2 instructional treatments. In ATI research, scholars usually assess participants in terms of individual difference factors such as working memory, aptitude, creativity, motivation, willingness to communicate or anxiety. Then, the effectiveness of some type of instructional treatment (e.g., task manipulation) is investigated in relation to the individual difference variable(s), involving either correlational designs (e.g., Fu & Li, 2019; Granena & Yilmaz, 2019; Nielson & DeKeyser, 2019; Révész, 2011) or comparison groups defined according to the learner variables (e.g., Yilmaz, 2013). Thus, ATI studies can be considered a subcategory of experimental research, and may also bear features of correlational/associate research.

To illustrate, Révész (2011) investigated whether three individual difference factors – linguistic self-confidence, anxiety, and self-perceived communicative competence – affect the extent to which L2 learners allocate attention to form-meaning connections during task-based interaction in a classroom context. Participants from six intact classes carried out a simple and complex version of the same type of decision-making task. Focus on form-meaning connections were captured in terms of a specific measure of speech production (use of conjoined clauses); global measures of complexity, accuracy, and fluency; and incidence of language-related episodes. Self-report questionnaires were used to elicit information about the participants with regard to the three individual difference variables. To assess the potential moderating effects of the individual difference factors, a series of correlational analyses were conducted for the simple and complex conditions separately.

Yilmaz (2013) provides a good example of a study where a comparison group design was adopted for some of the statistical analyses. This experiment examined the extent to which working memory capacity and language analytic ability influence the impact of two types of feedback on L2 development. Participants were assigned to three groups (explicit correction, recasts, and control), and received feedback according to their respective conditions during task-based work. Oral production, comprehension, and recognition tests were employed to assess changes in learners' knowledge of the target constructions. The operation span task and a subtest of LLAMA were used to measure working memory capacity and language analytic ability, respectively. To gauge whether the individual difference factors moderated the effectiveness of feedback types, Yilmaz first ran a series of mixed model ANCOVAs, with time as a within-subject variable, feedback group as a between-subjects factor, and working memory and language analytic ability as covariates. When the analyses yielded a significant ATI, the researcher converted the individual difference factor into a categorical variable, that is, divided

the participants into two groups: learners with scores above the median were regarded as high, and learners with scores below the median were considered low with regard to the aptitude factor. Then, some follow-up analyses were carried out involving aptitude as a categorical variable.

1.4 Descriptive, Non-experimental Research on Tasks
There has also been an increasing interest among TBLT researchers in conducting descriptive research that explores what happens during task-based interaction. Descriptive studies typically involve preparing audio-, video-, or screen-recordings of learners while they are engaged in pedagogic tasks. Then, researchers transcribe the recordings and analyse the data adopting an approach aligned with the theoretical orientation of their research and the focus of their research questions.

The aim of some descriptive research has been to capture task-based work in actual, unmanipulated classroom contexts. In this line of research, scholars usually utilise analytical frameworks such as interaction, multimodal, or conversation analysis, often inspired by a socio-cultural view of SLA. One focus of such studies has been to investigate how learners talk during task-based work, assuming an emic perspective. For example, Markee and Kunitz (2013) employed conversation analysis to study the interactional patterns of three Italian as a foreign language learners. The students were recorded during task work in their regular Italian language classes. The data comprised about three hours of video recordings collected over three weeks. From this dataset, four speech events were chosen for further analysis, each involving the planning stage prior to task performance. The researchers analysed the video transcripts in meticulous detail focusing on conversational features such as repair and turn-taking; types of embodied action, like gestures, body posture, and eye-gaze behaviours; and use of tools external to the task, including the computer and the notes that learners had taken while engaged in task-based planning.

Other researchers have taken an etic viewpoint when analysing natural task-based interaction in classroom settings, relying on pre-determined coding schemes, either adopted or adapted from previous research. For instance, researchers often code task-based interaction for negotiation of meaning and feedback episodes, categories which are derived from cognitive-interactionist approaches to SLA. Thus, Gurzynski-Weiss and Révész (2012) examined the extent to which the provision and immediate use of instructor feedback was related to whether the feedback occurred during tasks or non-tasks, unfocused or focused tasks, or the pre-, during-, or post-task stages. Twenty-three lessons were video-recorded from Spanish foreign language university courses. Next, the transcripts of the recordings were coded according to several interactional and task features based on a coding scheme that originated from theory and previous empirical findings.

In some descriptive classroom research, unlike in Markee and Kunitz (2013) and Gurzynski-Weiss and Révész (2012), where the interactions were naturally-occurring, the task-based materials have been developed by the classroom teacher and the researcher(s) together. The aim of this type of collaborative approach is to explore tasks that are of theoretical or practical interest to the researcher, but at the same time ensure that they remain aligned with normal classroom activities, and that the design thereby maintains ecological validity. For example, Mackey's (2002) study included data from three 50-minute lessons where ESL learners

completed task-based activities co-designed by the researcher and teacher. Participants also took part in a stimulated recall interview after the three lessons were over. The data analysis involved coding transcripts of oral interaction and stimulated recall comments in terms of interactional processes. A more recent study by Oliver, Philp and Duchesne (2017) also provides a good example of how task-based interaction can be explored in real classroom contexts through collaboration between teachers and researchers. As in Mackey (2002), the tasks used in the study were designed in cooperation with the classroom teachers. The dataset included transcriptions of interaction among children over five task-based sessions in their regular classroom context. The researchers adopted a bottom-up approach during the coding process by letting coding categories emerge from the data. Then, the resulting categories were labelled, informed by the existing literature on features of social interaction, task management, and cognitive involvement.

1.5 Case Studies of Teachers and Task-based Programmes
The case study is another approach to TBLT research. Case studies have been used to investigate how learners engage in task-based work, how teachers implement tasks, and how task-based programs work. The aim of this type of research has been to give a detailed picture of individual cases of learners, teachers, or programmes by describing them holistically and in depth in their own task-based environments. Case studies typically combine various data-collection methods and analytical approaches to capture the characteristics of a case or multiple cases in task-based contexts.

Research by Baba and Nitta (2014) and Nitta and Baba (2017) well exemplify how the case study approach can be used to investigate learners in task-based contexts. From a larger dataset, the researchers observed the longitudinal effects of task repetition on two students' writing development. The students engaged in repeating a writing task 30 times, once every week over a period of one academic year. Each time the participants completed a 10-minute writing output followed by reflective comments. Baba and Nitta (2014) focused on changes in students' writing fluency. Nitta and Baba (2017), in addition, analysed students' written outputs in terms of syntactic and lexical complexity, and considered their self-reflection from the perspective of self-regulation processes.

A seminal study by Samuda (2001) provides a good example of a case study considering the teacher's role in a task-based lesson. The researcher gathered audio and video-recordings of an English for Academic Purposes teacher and her class, as well as samples of the students' writing throughout a semester, spending a morning every week observing the class and the teacher. Using transcripts of classroom discourse, the study gives an in-depth description of one task-based lesson from beginning to end, exploring how the teacher gradually draws learners' attention to new language in the context of task-based interaction.

As more recent study by Andon (2018) also illustrates how a case study approach can be used to investigate teachers from the perspective of TBLT. The goal of this research was to explore the extent to which TBLT principles were represented in the practices and beliefs of three EFL teachers employed in UK private language school settings. Data collection involved the researcher observing lessons and conducting semi-structured ethnographic interviews with the teachers. The observations were carried out to gain information about the teachers' classroom practices and to gather a basis for comparing these with their perceived practices. The

observational data were also used for eliciting participants' views on specific activities that had taken place in the lessons. The aim of the interview schedule was to elicit participants' views on tasks and TBLT, but the researcher also allowed the teachers to raise issues, which were followed up on when considered relevant to the focus of the study. Andon adopted a data-driven, inductive approach to data analysis, but the process was also informed by the researcher's understanding of key characteristics and principles of task-based teaching.

The case study approach can also be employed to investigate task-based programs. Studies by McDonough and Chaikitmongkol (2007) and Carless (2004) provide good examples of this. Adopting a longitudinal design, McDonough and Chaikitmongkol (2007) aimed to investigate teachers' and learners' responses to a new task-based EFL program at a Thai university and to examine the ways in which any concerns raised by the teachers and students were handled in the program. During a 12-month period, the researchers collected data from multiple sources: learning notebooks, task and course evaluations, observations, field notes, and interviews. The data, oral and written, were subjected to qualitative, recursive analysis. In other words, the focus of the data collection was informed by the researchers' reflections on data that had been previously gathered. For example, the interview topics were guided by insights that had emerged from earlier class observations. Unlike McDonough and Chaikitmongkol, Carless (2004) primarily focused on teachers in his case study of a task-based program. This study employed a multiple case study approach to evaluate the behaviours and perspectives of three primary school teachers who were in the process of implementing a newly-introduced task-based program in Hong Kong. Carless observed the teachers in three cycles, each cycle entailing five to six classroom observations. The teachers' views were tapped through an attitude scale and interviews. The triangulation of these sources led to an understanding of how the implementation of the program was influenced by the beliefs of the teachers and practical issues inherent in their institutional contexts.

1.6 Practitioner Research
Practitioner research is another type which can be used to study task-based teaching and learning. Practitioner research, as its name suggests, is typically carried out by teachers in their own instructional settings. Two types of practitioner research that have been employed to investigate TBLT are action research and micro-evaluation of tasks.

Action research involves teachers, collaboratively or individually, in rounds of identifying, reflecting on, and finding solutions to problems that occur in their own specific task-based contexts. Given the cyclic nature of action research, the focus frequently develops as the investigation proceeds, with the teacher-researcher engaging in continuous revision of their TBLT practice. A good example of a TBLT action research project is a study by Shart (2008). The context was a beginner-level German class at a Japanese university, where Shart was the course instructor. Over a period of one year, Shart prepared weekly reflections on how the class was progressing and produced a thorough description of all the sessions he taught. In addition, another researcher, not involved in teaching the class, conducted focus group and individual interviews with the students, obtained students' perceptions about the classes through e-mail, and made classroom observations. The project was conducted in a number of stages. Shart first recognised the need for a language course that is aligned with the needs of his students. Drawing on his existing language teaching experience and understanding of the context, he decided that

TBLT would be a suitable pedagogical approach. In the next step, he designed a project to investigate the TBLT course he was going to teach. In the stages to follow, the teacher-researcher continued to refine the course, taking into account his own reflections and perceptions, the observations of the outside researcher, and the insights gained from the students' e-mails and interview comments. This cyclic approach proved helpful in reaching an improved understanding of the benefits and challenges entailed in implementing task-based teaching in the teacher-researcher's own pedagogic context.

Micro-evaluations of tasks are concerned with exploring whether a task works as intended (Ellis, 2011, 2015; Ellis et al., 2020). According to Ellis (2011), a possible procedure for evaluating tasks involves the following steps. First, the researcher needs to provide a thorough description of the task, which can later be used as a basis for the evaluation. Next, the aims of the evaluation should be determined; for example, whether the task succeeds in achieving the teacher's goals and whether it leads to unanticipated processes and outcomes. Data collection can start before the task (e.g., establishing what learners already know or can do), can take place during task performance (e.g., documenting how learners perform the task), and may continue after task performance (e.g., obtaining students' comments and perceptions about the task). Then, the researcher analyses the data, possibly through triangulation of various data sources. Based on the results of the analysis, the teacher-researcher can conclude whether the task was successful and what modification might need to be implemented to make it work better. Ellis (2015) describes a number of micro-evaluations of tasks, which were carried out by teachers as part of an MA-level TBLT course. The teachers followed the steps outlined in Ellis (2011), and took the form of what Ellis refers to as student-based and response-based evaluations. The teachers obtained student-based data largely through administering a short perception questionnaire to students. The response-based components involved the collection of either product- or process-based evidence. Investigation of the product was concerned with establishing whether the learners had achieved the intended task outcomes, whereas the process element examined the processes in which learners engaged during task performance (e.g., by looking at interactional patterns or task engagement).

Although there are many advantages to practitioner research, this type of design also has some limitations that need to be taken into account when interpreting the findings. One disadvantage is that the results often cannot be generalised to other contexts, given that the researchers typically develop action research plans to address their local problems or design tasks tailored to the particular characteristics of their students. Other issues specific to action research are that scholars cannot include control groups in their designs or control for extraneous factors inherent in classroom research. If such challenges cannot be overcome, the validity and reliability of the research will inevitably suffer, limiting the generalisability of findings. As Mackey and Gass (2015) note, for action research to be able offer insights for the broader community, it also needs to adhere to methodological standards accepted in the field. Nevertheless, in cases where this is not possible, the findings are still likely to prove interesting to fellow practitioners who work in similar contexts or need to deal with similar challenges (Mackey, 2017).

1.7 Systematic Research Syntheses
With the available research base growing, TBLT researchers increasingly use meta-analytic and synthetic techniques to summarize and review the results of empirical research on TBLT.

Systematic research syntheses, such as meta-analyses and narrative reviews, intend to find, analyze, and scrutinize primary studies carried out on a specific research topic. The principal aim of systematic research syntheses is to give a comprehensive summary of existing findings, research foci, and/or methodological approaches in the area studied.

Meta-analyses, a particular type of systematic research synthesis, can be employed to synthesize the results of quantitative studies by means of statistical analyses. So far, a number of meta-analyses have been conducted on TBLT-related topics, such as task-based interaction (Cobb, 2010; Keck et al., 2006), task complexity (Jackson & Suethanapornkul, 2013; Sasayama, Malicka, & Norris, 2015), and TBLT programs (Bryfonski & McKay, 2019). We will consider Jackson and Suethanapornkul (2013) in more detail to exemplify a TBLT research synthesis and meta-analysis. The authors set out to review previous empirical research on Robinson's (2001) Cognition Hypothesis, a framework proposed to model how task manipulations may affect L2 performance and development. The researchers focused on one prediction of the framework: when task complexity is increased along resource-directing dimensions, L2 production will be more complex and accurate but less fluent. First, the researchers conducted a comprehensive literature search, attempting to identify all studies that had investigated the Cognition Hypothesis before 2010. The authors found 47 studies with a focus relevant to the intended aims of the meta-analysis and synthesis. In the next step, they employed eight inclusion criteria to select studies for the synthesis, resulting in a pool of 17 published studies. These were synthesised taking account of key design features, including the task variables studied, the outcome measures used, the task conditions investigated, and the modalities of tasks in the research. Then, nine studies, with comparable aims and designs, were chosen to be included in a meta-analysis. Finally, for this set of studies, the researchers calculated combined effect sizes to examine the effects of increasing task complexity on syntactic complexity, lexis, and accuracy.

Rather than conducting a meta-analysis of previous TBLT research findings, Plonsky and Kim (2016) carried out a systematic review of the foci of studies exploring task-based learner production and the methodological features employed in this line of research. The authors first identified 85 primary studies investigating language production during task-based work, published between 2006 and 2015. Next, the studies were coded for their research focus (e.g., interactional features, CAF measures), contextual factors (e.g., laboratory versus class, institutional setting), and demographic variables (e.g., proficiency, age). In addition, Plonsky and Kim categorised the studies in terms of a number of methodological characteristics related to their design, the sampling and analytical procedures employed, and the level of transparency in reporting. Drawing on the results, the researchers put forward a number of suggestions for future TBLT research.

A qualitative research synthesis is a third type of systematic review that has been used to summarise and critique previous TBLT research. Chong and Reinders (2020) employed this approach to synthesise previous qualitative research on technology-mediated TBLT published between 2002 and 2017. Adopting a grounded theory approach, the authors synthesised the data obtained from 16 primary studies that utilised either qualitative or mixed methods designs. In the case of mixed-methods studies, the researchers only included the qualitative findings in the synthesis. While staying open to themes emerging from the data, the authors were interested in identifying themes with regard to the characteristics, opportunities provided by and limitations of

technology-mediated tasks. Relying on the qualitative software NVivo, the researches created 332 initial codes, which generated four conceptual, 10 descriptive, and 31 sub-categories. In addition to the topic they pre-specified, the data also yielded insights into what factors influence the effectiveness of technology-mediated TBLT.

Conducting meta-analyses and other types of systematic reviews are clearly important for the field of TBLT, as they can offer recommendations for teachers based on the aggravated results of many studies on a TBLT-related issue. However, as Sato and Loewen (2019, p. 13) note, given that instructed SLA is a relatively new field, researchers often examine new factors and techniques; thus, narrative reviews are likely to be comparably useful for teachers, as these can provide them with information about new techniques that they could trial in their own practice.

2. Issues in TBLT Research Methodology and Suggestions for Further Research

Having reviewed key methods that have been employed to study the role of tasks in second language teaching, a discussion follows of current issues in TBLT research methods. Also considered is how some of the methodological challenges might be overcome in future research.

2.1 Addressing Tensions Between Internal and Ecological Validity

As in other areas of instructed SLA research, a key challenge for TBLT researchers is to strike a balance between internal and ecological validity. While internal validity is concerned with the soundness of the design of empirical research, ecological validity has to do with the extent to which the research findings can be extended to real TBLT settings. Arguably, there is a need to conduct tightly controlled TBLT experiments, as these can help isolate variables that might affect task-based performance and development. However, the danger is that, due to the careful control for potential confounding factors, experimental studies become so artificial and removed from actual classrooms that the findings no longer seem to have implications for actual TBLT practice. To minimise this risk, researchers could start by observing the current practices and learner behaviours in the type of task-based settings for which they would like to draw implications. Then, the observations made could inform the development of the materials and procedures in subsequent experiments (Lightbown & Spada, 2019; Rogers & Révész, 2020). Ecological validity can also be enhanced through collaboration with teachers when developing tasks, task manipulations, and task-based lessons. As mentioned previously, a few TBLT studies have successfully adopted this approach (e.g., Kim, 2012; Mackey, 2002; Oliver at al., 2017). Finally, another way to deal with potential threat to ecological validity is to employ quasi-experimental rather than true experimental designs (Sato & Loewen, 2019). Given that quasi-experimental research often takes place in classrooms, it is likely to have greater potential for informing pedagogy. When conducting classroom studies, however, researchers need to make sure that they minimise the disruption of classroom activities, do their best to maintain objectivity, and comply with ethical issues pertinent to classroom research (Mackey, 2017).

2.2 Need for More Developmental and Longitudinal Research

Similar to other subfields of instructed SLA research, there is a lack of longitudinal studies on TBLT. Although the past two decades have seen a growth of studies investigating task-based development, most of the developmental research is still short-term, usually spanning not longer than two to four weeks. Also, these studies, the majority focusing on the effects of engaging in

task-based interaction (see Cobb, 2010; Keck et al., 2006; Mackey & Goo, 2007 for meta-analyses), typically had a narrow focus, investigating the acquisition of specific linguistic features rather than improvement in global proficiency. One could argue that, to inform and guide TBLT pedagogy, it would be necessary to conduct studies that take academic terms and even years, gauging overall L2 development in actual TBLT settings. However, the issue with such long-term studies is that they "tend to (and perhaps must) prioritize ecological validity over predictive validity" (Ellis et al., 2020, p. 300). Over extended periods of time, it is challenging to control for the large array of extraneous factors that can potentially affect classroom learning. Another practical problem is that carrying out longitudinal studies is highly labour-intensive, requiring a lot of researcher time and strong institutional commitment. These are challenging to secure in most contexts, due to low availability of research funding and already high demands on teachers. In light of this, it would appear more realistic for researchers to strive to conduct longitudinal studies that last for shorter periods (e.g., six to ten weeks). Such studies will allow for observing development in specific areas of task-based performance, serving as useful stepping stones to establishing the longer-term effects of task-based learning and teaching.

2.3 Focus on Processes and Products
To date, TBLT research has primarily been concerned with the products of task-based use and learning, mainly employing outcome measures such as CAF or indices gauging the use or knowledge of specific linguistic features. For the purposes of theory construction and informing pedagogical practices, however, it is also important to examine the processes in which learners engage during task-based work (Révész, 2014). Process-oriented research is, for example, warranted to explore the cognitive processes in which learners engage when they perform tasks. As Révész (2019) reviewed, there are a number of techniques that TBLT researchers have at their disposal to examine task-generated cognitive processes, including subjective techniques (e.g., questionnaires, interviews, think-aloud and stimulated recall protocols), as well as more objective tools (e.g., dual-task methodology, keystroke-logging, screen-recording, eye-tracking, fMRI).

In addition to looking at task-based processes, it would also be beneficial for future studies to focus more on links between process- and product-based measures. While there has been an increased interest in process-product relationships in the larger field of instructed SLA (e.g., Godfroid, Boers, & Housen, 2013; Pellicer-Sánchez, 2016), relatively few studies have looked into them in the area of TBLT. Among the early examples are interactionist TBLT studies that have examined the extent to which the frequency of interactional features (e.g., Adams, 2007) and cognitive activities (e.g., Mackey, 2006) during task performance predict L2 development. More recently, a few researchers have also begun to investigate how task variables may affect relationships between task-generated cognitive processes and task-based performance and development (e.g., Kim et al., 2015; Révész et al., 2017).

2.4 Triangulation of Sources
While in some types of TBLT research, such as the case study paradigm, data triangulation is a core feature of research designs, this methodological practice has been less widespread in cognitively-oriented TBLT research. However, in recent years, as in instructed SLA research in general (King & Mackey, 2016; Mackey & Gass, 2016), there has been an increasing trend towards collecting and triangulating multiple data sources. The rationale for utilising designs with various data sources is that the combination of different data-collection techniques, due to

inherent limitations associated with each, is likely to yield more valid and complete insights than use of a single method (Révész, 2019). As discussed earlier, task complexity researchers increasingly rely on and triangulate multiple measures when providing independent evidence for the validity of their task manipulations to enhance the credibility of their validity argument. Révész, Michel and Gilabert (2016; Michel, Révész, & Gilabert, 2014) collected data through four methods - dual-task methodology, self-perception questionnaires, eye-tracking and stimulated recall - to tap the effects of task complexity manipulations on task-generated cognitive processes. Researchers have also combined verbal protocol data, such as the stimulated recall procedure with keystroke-logging (Charoenchaikorn, 2019; Révész et al., 2017; Révész et al., 2019), eye-tracking (Révész et al., 2019), Google docs (Stiefenhöfer & Michel, 2019) and screen-recordings (Charoenchaikorn, 2019), to study task-based L2 writing processes. In each of these studies, triangulating various methods, as expected, allowed the researchers to achieve richer and more valid conclusions. In light of this, more widespread use of data triangulation would appear to benefit cognitively-oriented TBLT research in the future.

2.5 Data Reporting and Transparency
In their methodological synthesis of research on task-based language production, Plonsky and Kim (2016) point to a number of problems in data reporting and make a series of recommendations that researchers should follow to improve reporting practices in quantitative TBLT research. For example, Plonsky and Kim found that not all studies reported and interpreted reliability statistics, and visual displays of data were often missing or were ineffective. Plonsky and Kim also called for more detailed reporting of descriptive statistics including confidence intervals and effect sizes. In addition to improving reporting practices, it is crucial that TBLT researchers, regardless of their methodological orientation, make it a practice to share their instruments and data in open-science platforms such as IRIS. This will help increase the transparency and replicability of TBLT research, while also facilitating the education of TBLT scholars.

Suggested Readings

DeCosta, P. I., Valmori, L., & Choi, I. (2017). Qualitative research methods. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of second language acquisition* (pp. 522-540). New York, NY: Routledge.

DeKeyser, R. (2019) (Ed.). Aptitude treatment interaction in second language learning [Special issue]. *Journal of Second Language Studies, 2*(2).

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford, UK: Oxford University Press.

Loewen, S., & Philp, J. (2011) Instructed second language acquisition. Mackey, A., & Gass, S. (Ed.), *Research methods in second language acquisition: A practical guide* (pp. 53-73). Malden, MA: Wiley-Blackwell.

Mackey, A. (2017). Classroom-based research. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of second language acquisition* (pp. 541-561). New York, NY: Routledge.

Mackey, A., & Gass, S. M. (2015). *Second language research: Methodology and design* (2nd Ed.). New York, NY: Routledge.

McKinley, J. & Rose, H. (Eds.) (2020). *The Routledge handbook of research methods in applied linguistics*. New York, NY: Routledge.

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

Révész, A. (2014). Towards a fuller assessment of cognitive models of task-based learning: Investigating task-generated cognitive demands and processes. *Applied Linguistics*, *35*, 87–92.

Révész, A., Michel, M., & Gilabert, R. (2016). Measuring cognitive task demands using dual task methodology, subjective self-ratings, and expert judgments: A Validation Study. *Studies in Second Language Acquisition*, *38*, 703-737.

Discussion Questions
1. What do you see as the main benefits and disadvantages of conducting TBLT research in classroom and laboratory settings?
2. What data sources would you ideally triangulate to investigate a TBLT topic of interest to you?
3. In your view, what research designs should researchers use more extensively to help reach valid conclusions about the effectiveness of TBLT?

References
Adams, R. (2007). Do second language learners benefit from interacting with each other? In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 29-51). Oxford: Oxford University Press.

Albert, Á., & Kormos, J. (2004). Creativity and narrative task performance: An exploratory study. *Language Learning, 54*, 277–310.

Andon, N. J. (2018). Optimal conditions for TBLT?: A case study of teachers' orientation to TBLT in the commercial EFL for adults sector in the UK. In V. Samuda, M. Bygate & K. Van den Branden (Eds.), *TBLT as a researched pedagogy* (pp. 132–164). Amsterdam: John Benjamins.

Baba, K., & Nitta, R. (2014). Phase transitions in development of writing fluency from a complex dynamic systems perspective. *Language Learning, 64*, 1–35.

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*, 689–725.

Bryfonski, L., & McKay, T. H. (2019) TBLT implementation and evaluation: A meta-analysis. *Language Teaching Research, 23*, 603–632

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21– 46). Amsterdam: John Benjamins.

Carless, D. (2004). Issues in teachers' reinterpretation of a task-based innovation in primary schools. *TESOL Quarterly, 38*, 639–662.

Charoenchaikorn, V. (2019). *L2 revision and post-task anticipation during text-based synchronous computer-mediated communication (SCMC) tasks.* Unpublished doctoral dissertation, Lancaster University, UK.

Chong S.W., & Reinders H. (2020). Technology-mediated task-based language teaching: A qualitative research synthesis. *Language Learning & Technology*, *XX*(X), XX–XX.

Cobb, M. (2010). *Meta-analysis of the effectiveness of task-based interaction in form-focused instruction of adult learners in foreign and second language teaching.* Unpublished doctoral dissertation, University of San Francisco, CA, USA.

De la Fuente, M. J. (2006). Classroom L2 vocabulary acquisition: Investigating the role of pedagogical tasks and form-focused instruction. *Language Teaching Research, 10*, 263–295.

Dörnyei, Z., & Kormos, J. (2000). The role of individual and social variables in oral task performance. *Language Teaching Research, 4*, 275–300.

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

Ellis, R. (2011). Macro- and micro-evaluations of task-based teaching. In B. Tomlinson (Ed.), *Materials development in language teaching* (pp. 212–235). Cambridge: Cambridge University Press.

Ellis, R. (2015). Teachers evaluating tasks. In M. Bygate (Ed.), *Domains and directions in the development of TBLT: A decade of plenaries from the international conference* (pp. 247–270). Amsterdam: John Benjamins.

Ellis, R. & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research.* London: Routledge.

Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2020). *Theory and practice of task-based language teaching*. Cambridge: Cambridge University Press.

Gilabert, R., Barón, J., & Llanes, A. (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*, 367–395.

Godfroid, A., Boers, F., & Housen, A. (2013). An Eye for Words. Gauging the role of attention in L2 vocabulary acquisition by means of eye tracking. *Studies in Second Language Acquisition, 35*, 483–517.

Gurzynski-Weiss, L., & Révész, A. (2012). Tasks, teacher feedback, and learner modified output in naturally occurring classroom interaction. *Language Learning, 62,* 851–879.

Housen, A., & Kuiken, F. (2009) Complexity, accuracy and fluency in second language acquisition. *Applied Linguistics, 30*, 461–473.

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.

Jarvis, S. (2013). Defining and measuring lexical diversity. In S. Jarvis & M. Daller (Eds.), *Vocabulary knowledge: Human ratings and automated measures* (pp. 13–45). Amsterdam: John Benjamins.

Keck, C., Iberri-Shea, G., Tracy, N., & Wa-Mbaleka, S. (2006). Investigating the empirical link between task-based interaction and acquisition: A meta-analysis. In Norris, J. M. & Ortega, L. (Eds.), *Synthesizing research on language learning and teaching* (pp. 91–131). Amsterdam: Benjamins.

Kim, Y. (2009). The effects of task complexity on learner-learner interaction. *System, 37*, 254–268.

Kim, Y. (2012). Task complexity, learning opportunities, and Korean EFL learners' question development. *Studies in Second Language Acquisition, 34*, 627–658.

Kim, Y., Payant, C., & Pearson, P. (2015). The intersection of task-based interaction, task complexity, and working memory: L2 question development through recasts in a laboratory setting. *Studies in Second Language Acquisition, 37,* 549–581.

King, K., & Mackey, A. (2016). Research methodology in second language studies: Trends, concerns and new directions. *The Modern Language Journal, 100(s)*, 209–227.

Kourtali, N., & Révész, A. (2020). The roles of recasts, task complexity, and aptitude in child second language development. *Language Learning, 70*, 179–218.

Lambert, C. P. & Engler, S. (2007). Information distribution and goal orientation in second Language task design. In M. P. Garcia-Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 27-43). Clevedon, UK: Multilingual Matters.

Lee, J. (2019). Task complexity, cognitive load, and L1 speech. *Applied Linguistics*, *40*, 506–539.

Lightbown, P., & Spada, N. (2019, March). *In it together: Teachers, researchers, and classroom SLA*. Plenary presented at the annual meeting of the American Association of Applied Linguistics. Atlanta, Georgia, USA.

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

Mackey, A. (2002). Beyond production: Learners' perceptions about interactional processes. *International Journal of Educational Research, 37,* 379–394.

Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics*, *27*, 405–430.

Mackey, A. (2012). *Input, interaction, and corrective feedback in L2 learning*. Oxford, UK: Oxford University Press.

Mackey, A. (2017). Classroom-based research. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of second language acquisition* (pp. 541–561). New York, NY: Routledge.

Mackey, A., Adams, R., Stafford, C. & Winke, P. (2010). Exploring the relationship between modified output and working memory capacity. *Language Learning, 60*, 501–533.

Mackey, A., & Gass, S. (2016). *Second language research: Methodology and design*. New York: Routledge.

Mackey, A. (2017). Classroom-based research. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 541–561). New York: Routledge.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In Mackey, A. (Ed.), *Conversational interaction in SLA: A collection of empirical studies* (pp. 408–452). New York, NY: Oxford University Press.

Malicka, A. & Levkina, M. (2012). Measuring task complexity: does L2 proficiency matter? In A. Shehadeh & C. Coombe (Eds.), *Task-based language teaching in foreign language contexts: Research and implementation* (pp. 43–66). Amsterdam: John Benjamins.

Markee, N., & Kunitz, S. (2013). Doing planning and task performance in second language acquisition: An ethnomethodological respecification. *Language Learning, 63*, 1–36.

McDonough, K., & Chaikitmongkol, W. (2007). Teachers' and learners' reactions to a task-based EFL course in Thailand. *TESOL Quarterly, 41*, 107–132.

McDonough, K., Crawford, W., & Mackey, A. (2015). Creativity and EFL learners' language use during a group decision-making task. *TESOL Quarterly, 49*, 188-198.

Michel, M. (2017). Complexity, accuracy, and fluency in L2 production. In S. Loewen &M. Masatoshi (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 66–84). New York: Routledge.

Michel, M., Révész, A., & Gilabert, R. (2014). Eye movement prompts in stimulated recall: tapping cognitive processes based on audio vs. visual stimuli. Paper presented at AILA, Brisbane, Australia.

Michel, M., Révész, A., Lu, X., Kourtali, N., & Borges, L. (in press). Investigating L2 writing processes across independent and integrated tasks: A mixed-methods study. *Second Language Research, X*, XX-XX.

Nielson, K. B., & DeKeyser, R. M. (2019). Working memory and planning time as predictors of fluency and accuracy. *Journal of Second Language Studies, 2*, 281–316.

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. 279-309). Amsterdam: John Benjamins.

Norris, J., & Ortega, L. (2003). Defining and measuring SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 717–761). Malden, MA: Blackwell.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30*, 555–578.

Nuevo, A. (2006). *Task complexity and interaction: L2 learning opportunities and development*. Unpublished doctoral dissertation, Georgetown University, Washington, DC.

Oliver, R., Philp, J. & Duchesne, S. (2017). Children working it out together: A comparison of younger and older learners collaborating in task based interaction. *System, 69*, 1–14.

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition, 38*, 97–130.

Plonsky, L., & Kim, Y. (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*, 437–470.

Révész, A. (2011). Task complexity, focus on L2 constructions, and individual differences: A classroom-based study. *Modern Language Journal*, *95*, 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,* 87–92.

Révész, A. (2019). Investigating task-generated cognitive processes: Methodological advances and challenges. Plenary talk presented at the Biennial International Conference on Task-Based Language Teaching (TBLT), Ottawa, Canada.

Révész, A., Kourtali, N., & Mazgutova, D. (2017). Effects of task complexity on L2 writing behaviors and linguistic complexity. *Language Learning, 67*, 208–241.

Révész, A., Michel, M., & Gilabert, R. (2016). Measuring cognitive task demands using dual task methodology, subjective self-ratings, and expert judgments: A Validation Study. *Studies in Second Language Acquisition*, *38*, 703–737.

Révész, A., Michel, M., & Lee, M. (2019). Exploring second language writers' pausing and revision behaviours: A mixed methods study. *Studies in Second Language Acquisition, 41*, 605–631.

Révész, A., Sachs, R., & Hama, M. (2014). The effects of task complexity and input frequency on the acquisition of the past counterfactual construction through recasts. *Language Learning, 64*, 615–650.

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* (pp. 287–318). New York: Cambridge University Press.

Rogers, J. & Révész, A. (2020). Experimental and quasi-experimental designs. In H. Rose & J. McKinley (Eds.), The Routledge handbook of research methods in applied Linguistics (pp. 133–143): New York: Routledge.

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. 119–140). London: Longman.

Sasayama, S. (2016). Is a 'complex' task really complex? Validating the assumption of cognitive task complexity. *The Modern Language Journal, 100*, 231–254.

Sasayama, S., Malicka, A., & Norris, J. (2015). Primary challenges in cognitive task complexity research: Results of a comprehensive research synthesis. Unpublished paper presented at the 6th Biennial International Conference on Task-Based Language Teaching (TBLT), Leuven, Belgium.

Sato, M., & Loewen, S. (2019). Towards evidence-based second language pedagogy: Research proposals and pedagogical recommendations. In M. Sato & S. Loewen (Eds.), *Evidence-based second language* pedagogy*: A collection of instructed second language acquisition studies* (pp. 1–23). New York, NY: Routledge.

Shart, M. (2008). What matters in TBLT: Task, teacher or team? An action research perspective from a beginning German language classroom. In J. Eckerth & S. Siekman (Eds.), *Task-based language learning and teaching* (pp. 47-66). Berlin, Germany: Peter Lang.

Shintani, N. (2013). The effect of focus on form and focus on forms instruction on the acquisition of productive knowledge of L2 vocabulary by young beginner learners. *TESOL Quarterly 47*, 36–62.

Shintani, N. (2015). The incidental grammar acquisition in focus on form and focus on forms instruction for young, beginner learners. *TESOL Quarterly 49*, 115–140.

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

Skehan, P. (2009) Modelling second language performance: Integrating complexity, accuracy, fluency and lexis. *Applied Linguistics 30*, 510–532.

Torres, J. (2018). The effects of task complexity on heritage and L2 Spanish development. *Canadian Modern Language Review, 74*, 128–152.

Wang, Q. (2019). Chinese EFL learners' motivation and anxiety in a task context and the effects of individual difference on task performance. Unpublished doctoral dissertation, University of Auckland, Auckland, New Zealand.

Yilmaz, Y. (2013). Relative effects of explicit and implicit feedback: The role of working memory capacity and language analytic ability. *Applied Linguistics, 34,* 344–368.

Zalbidea, J. (2017). One task fits all? The roles of task complexity, modality, and working memory capacity in L2 performance. *The Modern Language Journal*, *101*, 335–352.