CrossMark

Studies in Second Language Acquisition, **44** (2022), 483–506
doi:10.1017/S0272263121000310

Research Article

EXAMINING THE RELATIVE EFFECTS OF TASK COMPLEXITY AND COGNITIVE DEMANDS ON STUDENTS' WRITING IN A SECOND LANGUAGE

Ting Sophia Xu

Faculty of Education and Social Work, The University of Auckland

Lawrence Jun Zhang *

Faculty of Education and Social Work, The University of Auckland

Janet S. Gaffney

Faculty of Education and Social Work, The University of Auckland

Abstract

While many studies have investigated the effect of task complexity on L2 writing, little has been reported on the effects of intended task complexity manipulations on task-generated cognitive demands in L2 writing. This study, therefore, was designed to examine the relative effects of task complexity and cognitive demands on students' L2 writing. Two argumentative writing tasks were manipulated with varying numbers of elements and reasoning demands to be distinguished either as a simple or complex writing task. Self-ratings and dual-task methodology were adopted to validate the manipulations of task complexity. Thirty-one L2 learners, in the single-task group, were asked to complete two writing tasks and a post-task questionnaire. Participants in the dual-task conditions (30 in Experimental 1 and 31 in Experimental 2) were required to simultaneously complete the primary writing tasks and the secondary tasks. Results from self-ratings and dual-task experiments supported the efficacy of the task complexity manipulations.

This work was supported by a joint scholarship for doctoral study offered to the first author by The University of Auckland and the China Scholarship Council. We would like to thank the students for their participation in this research and Guoxing Chen, Ming Yang, and Lihui Song for their assistance in data collection. We also thank the editor-in-chief, Professor Susan Gass, the handling editor, Professor Andrea Révész, and the anonymous reviewers for their constructive suggestions that have helped improve the clarity of the manuscript.

*Correspondence concerning this article should be addressed to Lawrence Zhang, Faculty of Education & Social Work, University of Auckland, Private Bag 92601, Symonds Street, Auckland 1150, New Zealand. E-mail: lj.zhang@auckland.ac.nz

© *The Author(s), 2021. Published by Cambridge University Press.*

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

INTRODUCTION

With the development of theories of instructed second language acquisition (SLA) and pedagogic criteria for task design and sequencing (Long, 2015), an increasing number of studies have investigated the effect of task complexity on L2 writing (Cho, 2015; Kuiken & Vedder, 2019; Lee, 2020; Rahimi & Zhang, 2019; Révész et al., 2016; Sasayama, 2016). Task complexity is "the result of the attentional, memory, reasoning and other information processing demands imposed by the structure of the task on the language learner" (Robinson, 2001, p. 29). Different manipulations of task complexity can generate varying cognitive demands, differently affecting L2 writing performance (Robinson, 2007b, 2011a; Skehan, 1996, 1998; Zhang et al., 2017). Within the paradigm of task complexity, researchers have traditionally hypothesized, rather than validated, that the designed task complexity distinguished levels of cognitive demands (Baralt et al., 2014; Lee, 2019, 2020; Révész et al., 2016; Sasayama, 2016; Sasayama & Norris, 2019). For example, task complexity differential has been typically operationalized by designing simple and complex versions (Wang & Zhang, 2019). The complex version has usually been assumed to call for more cognitive demands than the simple one.

Independently measuring every construct, according to Norris and Ortega (2003), is vital for evaluating any theoretical framework. Although task complexity is the key independent variable in the task-based research, independently validating the difference in the manipulated complexity of task demands has received little attention. In a small number of studies that have measured the manipulations of task complexity, using a post-task self-assessment questionnaire has been the most commonly used method (Rahimi, 2019; Rahimi & Zhang, 2018; Robinson, 2001, 2007b). Additional cognitive-psychological measures to self-ratings have been called for to confirm presumed task complexity manipulations (Baralt et al., 2014; Housen & Simoens, 2016; Révész, 2014). Recently, only a few studies (Baralt, 2013; Donate, 2018; Lee, 2019; Révész et al., 2014, 2016; Sasayama, 2016) have made endeavors to examine the relative effects of task complexity and cognitive demands in L1/L2 oral tasks by using techniques such as dual-task methodology, time estimation, and eye-tracking. A paucity of task-based research, however, has provided separate validations of task complexity manipulations within L2 writing tasks by using multiple measurements (Lee, 2020), which may be one of the main reasons for the inconclusive findings of the effects of task complexity on L2 writing.

To fill the existing research gap, the lack of validation of task complexity in L2 writing, this study aimed to independently and multidimensionally validate the efficacy of the task complexity manipulations within writing tasks by using self-ratings and the dual-task methodology. Our study is expected to contribute to innovations in the independent assessment of intended task complexity of L2 writing by adopting multi-methods as a way of methodological triangulation. This is an initial effort to use the dual-task technique to measure the cognitive demands imposed by L2 writing tasks in task-based research.

LITERATURE REVIEW

TASK COMPLEXITY

Two well-known models of task complexity, Skehan's (1996, 1998) Limited Attentional Capacity Model and Robinson's (2001, 2005, 2011a) Cognition Hypothesis, have

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

identified a number of dimensions and variables by which task complexity is determined. Skehan (1996, 1998, 2018) distinguished three factors contributing to the complexity of tasks: code complexity, cognitive complexity, and communicative stress. These three dimensions influence the complexity of tasks and consequently, the amount of attention and cognitive demands the task requires of the learners (Skehan & Foster, 2001). Learners can attend only to one aspect of language production (i.e., complexity, accuracy, or fluency) when completing a cognitively demanding task, as "humans have a limited information processing capacity and must therefore prioritize where they allocate their attention" (Skehan, 2013, p. 189).

Contrary to Skehan's Limited Attentional Capacity Model, Robinson's (2001, 2005, 2011a, 2011b) Cognition Hypothesis holds that dimensions of task complexity belong to different attentional resource pools, namely, resource-directing dimensions that place cognitive/conceptual demands on learners, and resource-dispersing dimensions that impose performative/procedural demands on learners. Increases in task complexity along resource-directing dimensions, for example increasing the number of elements involved in the task, place cognitive demands on learners, which can direct learners' attentional resources to language form and facilitate "the development and acquisition of new L2 form-concept mappings" (Robinson, 2007a, p. 18).

Task complexity and cognitive demands are the key constructs in models of task complexity. A large part of existing empirical research has explored the effects of task complexity on L2 writing (Abrams, 2019; Ishikawa, 2007; Kang & Lee, 2019; Kuiken et al., 2005; Kuiken & Vedder, 2007, 2008, 2012; Ong & Zhang, 2010, 2013; Rahimi, 2019; Rahimi & Zhang, 2019; Révész et al., 2017). Little attention, however, has been paid to the validity of task complexity manipulations in L2 writing research. Therefore, this study was designed to explore ways in which the validity of intended task complexity manipulations can be assessed, that is, whether writing tasks designed to be more complex indeed generate higher cognitive demands during the writing process.

SELF-RATING SCALES AS MEASURES OF COGNITIVE DEMANDS

Subjective self-rating scales have been the most commonly used method to examine the presumed task complexity in the field of SLA (Révész et al., 2016; Sasayama, 2016).

As one of the earliest studies, Robinson (2001) independently examined the effects of task complexity on learner perceptions of task difficulty by using a 9-point self-rating scale. He found that participants rated a hypothesized-to-be simple task as significantly less difficult and less stressful than the designed complex task. To measure participants' perceptions of task difficulty, Robinson assessed their perceived levels of task difficulty, stress, confidence, interest, and motivation. The variables in Robinson's scale, however, do not have identical constructs with perceived cognitive demands, although they are often highly correlated (Brünken et al., 2010). Task difficulty refers to learners' perceptions of the cognitive demands arising from the task (Robinson, 2005, 2007a, 2011a), whereas cognitive demand reflects "the load that performing a particular task imposes on the cognitive system of a learner" (Paas & Van Merriënboer, 1994, p. 353). In the cognitive psychology research literature, mental effort is widely used as a direct measurement of cognitive demand, which reflects "the amount of resources actually allocated to accommodate the task demands" and "the amount of controlled processing in which the

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

individual is engaged" (Paas et al., 1994, p. 420). In contrast to task difficulty, which emphasizes learners' perceptions of the tasks, mental effort focuses on the actual effort made by individuals (Sasayama, 2016).

Although Robinson's task difficulty self-rating scale has been widely and actively used in recent task-complexity studies (Rahimi, 2019; Rahimi & Zhang, 2018, 2019; Robinson, 2007b), this scale does not exactly reflect cognitive demands placed on learners by the manipulations of task complexity. A multidimensional scale with a combination of mental effort and task difficulty is, therefore, desirable for providing a more comprehensive and scientific evaluation of perceived cognitive demands.

To date, some researchers (Donate, 2018; Lee, 2019, 2020; Révész et al., 2016; Sasayama, 2016) have used a scale that includes mental effort and task difficulty to tap into learners' cognitive demands in the field of task-based research. A two-item, paper-based questionnaire was first used by Sasayama (2016) to measure both task difficulty and mental effort: "(a) Was telling the story in English easy? (b) How much brain power did you use to complete this task?" (p. 238). She found that among the four oral tasks, the hypothesized simplest task was rated as the least difficult and the least mental-effort-demanding, while the task designed to be the most complex was rated as the most difficult, requiring the greatest mental effort. These may not, however, accurately represent the cognitive demands yielded by the four oral tasks per se, as participants completed each task under a dual-task condition. The results obtained for task difficulty and mental effort, therefore, may include not only participants' perceptions on performing the oral task but also the perceptions related to the dual-task condition.

Donate (2018) extended Sasayama's (2016) research on English learners of Spanish. She used the same subjective measures, self-ratings of task difficulty and mental effort and time estimation, to validate the cognitive demands posed by two oral narrative tasks, which were proven to be the simplest and the most complex tasks in Sasayama's (2016) study. Findings showed that the hypothesized-to-be complex task required significantly greater mental effort than the simple task. No significant differences, however, were found for self-perceived task difficulty and time estimation, although the intended complexity level of these two tasks had been confirmed in Sasayama's study. Donate (2018) speculated that the low level of proficiency for the novice learners of Spanish and the type of task might be factors that led to the unexpected non-significant results.

In Lee's (2019) study, a questionnaire with 9-point Likert scales on task difficulty, mental effort, and stress was used to measure native English-speakers' perceptions of task-generated cognitive demands. Three types of oral tasks (i.e., direction-giving map task, seating arrangement task, and car accident task) were employed with three intended versions of task complexity in each task-type: least-complex, mid-complex, and most-complex. The predictions that increased task complexity yielded more cognitive demands were borne out by the results of self-ratings on task difficulty, mental effort, and stress, although mental effort and stress ratings on the most and mid-complex versions of the map task did not differ significantly. Similar to Sasayama's (2016) research, all participants in Lee's study engaged in the dual-task condition. Consequently, the self-rating data may exaggerate the effects of task complexity manipulations.

Révész et al. (2016) used a 6-item computer-based questionnaire on a 9-point scale to measure the manipulations of the designed simple and complex versions of three pairs of oral tasks: a pair of picture narrative tasks, a pair of map tasks, and a pair of decision-making

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

tasks. In contrast to Sasayama's (2016) and Lee's (2019) research designs, participants in Révész et al.'s (2016) study were allocated into two groups: a dual-task condition and a single condition (without an added secondary task). Only students in the single condition completed the scales after each oral task. The findings showed that for the map and decision-making tasks, but not the narrative tasks, the hypothesized complex version was rated as more difficult and consumed more mental effort than the designed simple task version.

Lee (2020) adapted a multidimensional questionnaire from Lee's (2019) study, in addition to the measures of experts' judgments and time-on-task, to investigate the manipulations of task complexity in L2 writing in a Korean context. Task complexity within two decision-making writing tasks, a Hotel Task and a Venue Task, was manipulated by varying the number of elements as well as a variable of [± open solution]. Participants found the hypothesized-to-be complex task more difficult, stressful, and mental-effort consuming, and they were more pressed for time when planning and writing the designed complex task than the simple one. These results were consistent with findings shown by expert judgments and time-on-task measures. No objective techniques, such as dual-task method, for cognitive demands validation were employed to triangulate the findings obtained by subjective methods in this study.

Although prior studies (Donate, 2018; Lee, 2019, 2020; Révész et al., 2016; Sasayama, 2016) found that the multifaceted self-assessment is a viable way to tap into the task-generated cognitive demands, learners' self-perceptions may vary according to their proficiency (Donate, 2018; Sasayama, 2016), task type (Lee, 2019; Révész et al., 2016), and task modality. More research in the field of L2 writing is, therefore, warranted to test the utility of self-rating scales in confirming the manipulations of intended task complexity, especially within an argumentative task-type that has not been sufficiently validated.

DUAL-TASK METHODOLOGY AS A MEASURE OF COGNITIVE DEMANDS

Dual-task methodology, as suggested by its name, involves performing an additional secondary task simultaneously with a primary task. During the primary task, participants react to a variable-timed probe stimulus (i.e., secondary task) as quickly as possible. Dual-task methodology, as an objective measurement, serves as a direct measure of cognitive demands without the influence of mediating variables, and is independent from the learners' self-report (Brünken et al., 2003; Schoor et al., 2012).

The principle underlying the dual-task technique is that the amount of cognitive demand induced by the primary task will be reflected in the secondary task performance, which is normally assessed in terms of reaction time and accuracy rate (Brünken et al., 2003; Révész, 2014; Révész et al., 2014, 2016; Schoor et al., 2012). For instance, when participants completed a primary task with more cognitive demands, less accuracy and more reaction time were expected on their secondary task performance than when they performed a simple primary task. According to the suggestions of several researchers (e.g., Brünken et al., 2003; Marcus et al., 1996; Schoor et al., 2012), secondary tasks should be in the form of simple activities, such as an easy visual or auditory stimulus; provide as little disturbance as possible to the primary task; be continuously scored, independent of the primary task performance; and not suppress the quality or quantity of primary task performance.

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

Dual-task methodology has been employed in a great number of studies in the field of cognitive psychology (Brünken et al., 2003, 2010; Marcus et al., 1996; Schoor et al., 2012), and is considered a reliable and sensitive approach for measuring cognitive demands. Dual-task methodology, however, has been given insufficient attention in the field of SLA (cf. Révész et al., 2014, 2016; Sasayama, 2016).

Researchers have recently adopted dual-task methodology as a measure of task-generated cognitive demands in studies of task complexity. For example, Révész et al. (2014, 2016) conducted studies using it and other techniques to independently assess the validity of task complexity. In Révész et al.'s (2014) study, two oral tasks with different reasoning demands were the primary tasks in which participants were asked to describe a famous person's life by using a past counterfactual statement based on two past events they chose. The secondary task was to respond to the change of the computer background screen, which was programmed to randomly change the color to red or green during each 2,500-ms interval. Half the participants were instructed to respond to green (ignoring red), while the other half were asked to respond to red (ignoring green). Accuracy and reaction time were recorded to reflect the level of cognitive demands imposed by the primary tasks. The findings showed that participants produced higher accuracy rates on the secondary task when performing the designed simple primary task, although no significant differences were found in the secondary-task reaction time.

Révész et al. (2016) triangulated the findings obtained from the dual-task methodology and other techniques (i.e., self-rating scale and expert judgment) to respectively measure the cognitive demands imposed by three pairs of oral tasks as the primary tasks. The secondary tasks used in these two studies (Révész et al., 2014, 2016) were identical, although the primary tasks were quite different. Three pairs of designed simple and complex oral tasks were used as the primary tasks, and the three modes were picture narrative, map, and decision-making tasks. Participants achieved a higher mean accuracy rate but non-salient reaction time when carrying out the simpler version of the primary tasks. The results for accuracy rates were found to be consistent with the findings of self-ratings and expert judgments. Based on the findings for accuracy and reaction time in the two studies mentioned, Révész et al. (2016) postulated that accuracy rates might be more sensitive than reaction time under the color detection secondary-task condition.

Sasayama (2016) employed the dual-task method, together with time estimation and self-rating techniques, to measure intended task complexity. She selected a color-detection design as the secondary task. The color of letter A was programmed to change unpredictably from black to red and from red to black. The red color appeared during each 5,000-ms interval, while the duration of the black letter was set with a variation from 3,000 to 60,000 ms. Participants were required to react as quickly as possible only when letter A was changed from red to black. Four narrative oral tasks, which involved an increasing number of elements, were designed to vary the cognitive demands of the primary tasks. Sasayama found a trend that participants responded to the color stimulus the fastest when carrying out the hypothesized-to-be simplest task, and the slowest when engaging in the most complex one, although this difference was not significant. The same pattern, however, was not identified for the hypothesized intermediate complex versions. Sasayama (2016) reported consistent patterns of results on three measurements of task complexity: dual-task methodology, self-ratings, and time estimation.

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

Lee (2019) employed the dual-task method to validate three-levels of task complexity manipulations of three types of oral tasks completed by 42 English native speakers. For the secondary task in her study, participants reacted to screen color changes of a laptop, which changed at a 2,500 ms-interval from either white to red or white to green. Participants were instructed to press, as quickly and accurately as possible, the right shift key when the screen was changed from white to red, and the left shift key when changed from white to green. The findings showed that participants reacted slowest to the hypothesized most-complex versions than to the mid- and least-complex versions, which is consistent with the results on self-ratings, although no significant differences were found in accuracy. The task-complexity effects on secondary-task reaction time were also found to be moderated by the types of tasks. The results of this study may, however, not reveal the effects of intended task complexity on the cognitive demands of L2 learners. Compared to the native speakers, L2 learners who had language barriers may differentially react to the secondary-task color stimuli, perceive the difficulty of the tasks, and make distinct actual mental efforts during the (primary) task completion.

Révész et al. (2014, 2016), Sasayama (2016), and Lee (2019) pioneered the dual-task paradigm to independently measure cognitive demands of oral tasks in task-based research. In contrast to the result patterns found in Révész et al.'s studies (Révész et al., 2014, 2016), Lee (2019) found a stronger effect in the dual-task experiment for reaction time than for accuracy. Lee (2019) conjectured that the secondary task she used may have created an interference: The counterintuitive key-color assignments "may have been complex enough to capture different reaction times according to the complexity of the primary tasks" (p. 533). The different result patterns of reaction time and accuracy found in these three studies show that the distinct design of the secondary task might lead to participants' different responses to the stimulus. Given that employing the dual-task method to validate task complexity manipulations in the SLA field is still at a fledging period, inclusion of various versions of secondary tasks to measure the cognitive demands imposed by the same primary tasks is warranted. Additionally, the primary tasks in these three studies were oral, and the secondary tasks were visually presented. Some scholars commented that the secondary tasks should be chosen based on the modality of the primary tasks (e.g., Schoor et al., 2012). The visual secondary task with a color-detection design is obviously not appropriate when the primary tasks are written.

Reed et al.'s (1985) study is one of the few that have used dual-task methodology in the field of language writing research. Their study explored the effects of writing ability and mode of discourse on cognitive capacity engagement by employing dual-task methodology. Three writing tasks in forms of descriptive, narrative, and persuasive writing served as the primary tasks. The secondary task was to react to a tone at random from a microcomputer by pressing the space key. The tone was presented 20 times in total, randomly appearing 5 times at the prewriting stage, 10 times at the writing stage, and 5 times at the rewriting stage. Cognitive-capacity engagement was analyzed by participants' reaction time. The results showed that writers' cognitive-capacity engagement across different modes was differentially affected by their writing ability. Writers with advanced writing ability were most engaged when carrying out persuasive writing tasks and least engaged while writing descriptive essays; whereas writers with average writing ability were the most engaged when writing descriptive essays and least engaged in narrative writing tasks. Although not related to task complexity, Reed et al.'s (1985) study

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

provides a different viewpoint on using dual-task methodology in the field of writing research. The dual-task technique has rarely been used to measure the cognitive demands imposed on learners by writing tasks in the field of task-based research. The auditory stimulus and the sound-response format, as in Reed et al.'s (1985) research, can be applied to the secondary task to measure the cognitive demands of primary writing tasks on learners.

The literature review signals scant validation of manipulations of task complexity in L2 writing using multiple measures, although the effects of task complexity on EFL/L2 writing have been given increased attention. As the key independent variable in task-based research, the efficacy of task complexity manipulations in EFL/L2 writing is of vital importance and, therefore, needs to be validated.

RESEARCH QUESTIONS

This study addressed the following research questions:

1. Do different levels of intended task complexity result in different levels of cognitive demands during task performance, as measured by self-perceived task difficulty and mental effort self-assessment?
2. Do different levels of intended task complexity result in different levels of cognitive demands during task performance, as measured by dual-task methodology?

METHODS

PARTICIPANTS

Ninety-two learners of English in three intact classes from a university in China voluntarily participated in this study. Their mean ages were 20, with a range of 19–22, and the majority of them were female (90% female and 10% male). No participants had overseas learning experiences. Their years of learning English ranged from 8 to 14 years, with an average of 10.88 years.

Students in three intact classes as a whole were randomly assigned into three groups with different experimental conditions. Based on their last semester final exam results of *Integrated English Course* and *English Writing Course*, no significant differences were found in students' English proficiency across these three intact classes. One class with 31 students was assigned as a single-task group and the other two classes as dual-task groups, with 30 students in dual-task Experimental 1 and 31 students in dual-task Experimental 2.

INSTRUMENTS

Self-Perceived Task Difficulty and Mental Effort Questionnaire

The self-perceived *Task Difficulty and Mental Effort Questionnaire* was adopted from Révész et al. (2016). The paper-based self-rating scale, translated into Chinese, included six items with a 9-point response scale ranging from 1 (very easy/not at all) to 9 (very difficult/intensively). Based on the precedent of Révész et al. (2016), only two items (Item

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

1 and Item 3) out of six relevant to measuring cognitive demands are discussed in this study. Item 1 assessed participants' perceptions of mental effort and item 3 was about participants' feelings of the task difficulty.

Primary Writing Tasks

The two writing tasks were adapted from Cho's (2015) work with the topic "choosing the best roommates." The prompts of the two writing tasks were translated into Chinese and modified to suit the L2 learning context in China to ensure that all the participants would fully and quickly understand the topics and instructions. These tasks were chosen because students living in the dormitory on the university campus in China were familiar with and interested in this topic.

In the two argumentative writing tasks, task complexity was manipulated by varying the number of argument elements and reasoning demands. The designed-to-be simple task required learners to choose one pair of best roommates from four candidates, each marked by four attributes: personality, hobbies, sleeping patterns, and studying style. The designed complex task asked students to choose two pairs of best roommates from six candidates, each marked by six properties: personality, hobbies, sleeping patterns, studying style, favorite subjects, and individual sanitary habits. The complex task, compared to the simple version, was characterized by more argument elements and more reasoning demands, with two more candidates and two more properties of each candidate from which to choose. The rationale for such intended task complexity manipulations was that to complete the tasks with more argument elements and reasoning demands was expected to consume more cognitive demands than those with fewer elements and lower demands on reasoning (Cho, 2015; Kuiken & Vedder, 2007; Rahimi, 2019; Robinson, 2005, 2007b).

Secondary Tasks in Dual-Task Experiments

The secondary tasks in the dual-task experiments were simple, continuous auditory tasks. The two versions of secondary tasks were designed based on previous studies, with the format of the stimuli, such as stimulus-detection and stimulus-response, and frequency of the stimulus being taken into account (Brünken et al., 2003; Marcus et al., 1996; Reed et al., 1985; Révész, 2014; Révész et al., 2014; Sasayama, 2016; Schoor et al., 2012). Two dual-task experiments with different secondary tasks, a sound-detection version and a sound-response version, were involved, as different result patterns for reaction time and accuracy were found, likely due to distinct secondary-task designs in previous studies (Lee, 2019; Révész et al., 2014, 2016; Sasayama, 2016). Using two versions of secondary tasks to validate the intended task complexity manipulations of the same two primary tasks with similar participants is also a way to triangulate the findings obtained from dual-task methods in L2 writing.

Secondary Task 1. In dual-task Experiment 1, stimuli of the secondary task were two sounds, a high-pitch sound *di* (1,000 Hz) and a low-pitch sound *du* (300 Hz). The computer was programmed to generate *di* and *du* for 150 ms at random intervals in the range of 30–60 s. Participants were instructed to press "Z" on the keyboard as fast and accurately as possible, when they detected the sound through an earphone. Half the participants were required to react to *di* ignoring *du*, and half to *du* ignoring *di.* The reaction time and accuracy rates were obtained from E-prime 2.0 in milliseconds.

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

Secondary Task 2. In dual-task Experiment 2, the sound *di* (1,000 Hz) was programmed to occur for 150 ms in a variable interval between 20 and 40 s. Participants, upon hearing *di*, were asked to react as fast and accurately as possible to the stimulus. Participant's reaction time and accuracy in responses to *di* by pressing "Z" on the keyboard were recorded by E-prime 2.0 in milliseconds.

PROCEDURES

Participants in the single-task group were asked to complete two writing tasks, each task followed by the *Task Difficulty and Mental Effort Questionnaire.* Forty minutes was provided for each writing task and 5 min for the *Task Difficulty and Mental Effort Questionnaire.* A 2-week interval occurred between the two writing tasks, and the two tasks were counterbalanced to avoid practice and fatigue effects.

Participants in the dual-task conditions were required to complete the primary writing tasks and secondary tasks simultaneously. Participants were tested individually in groups up to 16 students at one time in the school computer room. Procedures for the two experiments were the same, except for the secondary tasks. Each experiment was conducted twice by the researcher with the help of two research assistants. The first time took about 90 min for each participant, including a practice phase, a baseline test, and the first dual-task test, and the second time took about 60 min, involving a warm-up phase and the second dual-task test. There was a 2-week interval between the two experiments, and the order in which the participants carried out the two primary writing tasks was counterbalanced (see Table 1). Task instructions were presented in Chinese to ensure ease of comprehension by participants.

The 15-min practice phase was provided to familiarize participants with task instructions and experimental procedures in the dual-task experiment. Participants were asked to write a self-introduction on the paper, and, at the same time, they reacted to the secondary task. During this phase, participants were encouraged to ask the researcher and research assistants any questions regarding the procedures. The researcher also monitored the procedures to ensure that participants were appropriately following the task instructions. Some students were given additional instructions as needed. No participants reported they had hearing problems or had difficulty in recognizing the sound(s).

TABLE 1. Writing-task counterbalance in the three groups

	Participants	First time	Second time
Single-task group	Half ($n = 15$)	T1	T2
	Half ($n = 16$)	T2	T1
Dual-task group (*di* & *du*)	Half ($n = 15$)	T1	T2
	Half ($n = 15$)	T2	T1
Dual-task group (*di*)	Half ($n = 16$)	T1	T2
	Half ($n = 15$)	T2	T1

Note: The tag of the two writing tasks, such as "simple" and "complex" was anonymized as task 1 (T1) and task 2 (T2). In dual-task groups, writing tasks were completed as the primary tasks.

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

For the baseline test, participants were asked to complete the secondary task in 20 min without an added primary writing task. The secondary task that participants performed during the baseline test was the same as what they would complete during the actual dual-task experiments. Participants in the two experiments were each instructed to respond to the sound stimulus as fast and accurately as possible, and they were also informed that their reaction time and accuracy rate would be recorded. After this phase, participants moved on to the 40-min actual dual-task experiment.[1] They were informed that the writing tasks were a test of their writing ability and would be rated in terms of content, organization, vocabulary, language, and mechanics to ensure that they could engage in the primary tasks during the dual-task experiments.[2] They were not allowed to consult any papers, books, dictionaries, or other materials during the experiment.

After a 2-week interval, participants engaged in the second experiment. While participants completed a warm-up exercise to become acquainted with the dual-task procedure, they engaged in the actual dual-task experiment.

DATA ANALYSES

Analysis of Dual-Task Experiments

Participants' responses to the secondary task were recorded every time they pressed the "Z" in keyboard to indicate their response to the sound stimulus. Their response data were analyzed by reaction time of the correct responses and the accuracy rate, following previous research (Révész et al., 2016).

Experiment 1. Accuracy was calculated by dividing the number of times participants correctly reacted to *di* and ignored *du* (half of them correctly reacted to *du* and ignored *di*) by the total number of stimuli (both *di* and *du*) occurring. Reaction time was considered for correct responses only.

Experiment 2. Accuracy rate was calculated by dividing the number of times participants correctly reacted to *di* by the total number of occurrences of *di.* Reaction time was calculated by dividing the summed reaction time of accurate responses to *di* by the total number of correct responses to *di.*

Statistical Analysis

The analyses for self-ratings were conducted based on single-task performances only. To examine the effects of intended task complexity on self-ratings, a series of paired-samples *t*-tests was applied. For dual-task experiments, one-way repeated-measures ANOVAs were conducted, with task conditions (i.e., the baseline condition, the simple primary task condition, and the complex primary task condition) as the repeated condition, and reaction time and accuracy rate as the dependent variables.

The alpha level was set at $p < .05$. η_p^2 for ANOVA, and Cohen's *d* for the *t*-test were employed to measure effect sizes. According to Cohen (1988), η_p^2 values of .01, .06, and .14, and *d* values of .20, .50, and .80, respectively, represent small, medium, and large effect sizes.[3]

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

RESULTS

TASK DIFFICULTY AND MENTAL EFFORT QUESTIONNAIRE

All data were first inspected for missing values and outliers with a cut-off set at three standard deviations from the mean. No missing values and no outliers were found in the data of the single-task group. Table 2 displays the descriptive results for the *Task Difficulty and Mental Effort Questionnaire* in the simple and the complex task versions. As Table 2 shows, the hypothesized complex task was rated, on average, to be more difficult and more mental-effort consuming than the task designed to be simple.

As the assumption of normality was met, separate paired-samples *t*-tests were computed for self-ratings of task difficulty and mental effort to examine whether there were significant differences between the simple and complex tasks. As expected, participants perceived the hypothesized complex task ($M = 5.61$, $SD = 1.56$) as requiring significantly more mental effort: $t\ (30) = -2.39$, $p = .024$, $d = -.42$, compared with the task designed to be simple ($M = 5.00$, $SD = 1.34$), and the effect size was medium. The same was true for participants' self-ratings on task difficulty for the presumed simple ($M = 4.71$, $SD = 1.13$) versus complex tasks ($M = 5.35$, $SD = 1.33$), $t\ (30) = -2.48$, $p = .019$, $d = -.52$. The effect size was medium. These findings indicate that the hypothesized-to-be complex task was more complex than the simple task in terms of self-perceived task difficulty and mental effort.

Pearson's correlation coefficients were also calculated, results indicating a strong but not perfect correlation between perceived task difficulty and mental effort (simple task: $r = .55$, $p = .001$; complex task: $r = .61$, $p < .001$). These results suggested that the two variables were related but not isomorphic (Brünken et al., 2010; Révész et al., 2016).

DUAL-TASK EXPERIMENT 1

Table 3 summarizes the descriptive statistics for reaction time and accuracy rate on the secondary task (*di* and *du* version) under the baseline, simple, and complex task conditions. Participants responded to the sound stimuli fastest under the baseline condition and the slowest when engaging in the hypothesized-to-be complex task. The accuracy rate of the participants' reaction decreased, as the task condition became more difficult.

The assumptions of normality and sphericity were examined and found to be satisfied, therefore, a series of one-way repeated-measures ANOVA was conducted to explore the

TABLE 2. Descriptive statistics for self-ratings of task difficulty and mental effort in the simple and complex tasks

		Simple task			Complex task		
	N	*M*	*SD*	95% CI	*M*	*SD*	95% CI
ME (Item 1)	31	5.00	1.34	[4.51, 5.49]	5.61	1.56	[5.04, 6.19]
TD (Item 3)	31	4.71	1.13	[4.29, 5.12]	5.35	1.33	[4.87, 5.84]

Note: The values represent ratings on a 9-point scale (1 = very easy/not at all, 9 = very difficult/intensively). ME = mental effort, TD = task difficulty.

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

effects of intended task complexity on participants' secondary task performance in terms of reaction time and accuracy. The ANOVA tests showed that each of the two analyses yielded a significant and large effect on participants' performance in the secondary task across three task conditions, reaction time: $F\,(2, 58) = 37.89$, $p < .001$, $\eta_p^2 = .57$; and accuracy rates: $F\,(2, 58) = 21.14$, $p < .001$, $\eta_p^2 = .42$. These results suggest that there were significant differences in reaction time and accuracy, across baseline, simple, and complex conditions.

Pairwise comparisons with Bonferroni adjustment were, therefore, performed to investigate where the statistically significant differences were located. For reaction time, participants reacted to the secondary task significantly faster under the baseline condition ($M = 543.91$, $SD = 123.12$) than either the simple condition ($M = 848.13$, $SD = 189.61$): $p < .001$, $d = -1.90$, or the complex condition ($M = 934.77$, $SD = 311.20$): $p < .001$, $d = -1.56$; the effect sizes were large. In contrast, intended increases in task complexity of the primary writing task did not have a significant impact on participants' reaction time, $p = .193$, $d = -.31$. These results indicate that the design of the dual-task methodology was sensitive, in terms of reaction time, to distinguish the baseline and experimental conditions. Increases in task complexity of the primary tasks, however, did not statistically affect participants' secondary-task reaction time.

Significantly higher accuracy rates were produced on the secondary task when participants were under the presumed simple primary task condition ($M = 94.19$, $SD = 4.26$), as compared with the complex primary task ($M = 91.60$, $SD = 4.62$), $p = .038$, $d = .58$. The effect size was medium. Participants under the baseline condition ($M = 98.11$, $SD = 2.17$) substantially outperformed those under either the simple primary task condition, $p = .002$, $d = 1.19$, or the complex condition, $p < .001$, $d = 1.86$; the effect sizes were large. The results reveal that accuracy as a measure of students' cognitive demands can sensitively separate the baseline and experimental conditions and predict participants' greater consumption of cognitive demands when the task complexity of the primary tasks was increased as intended.

In sum, significant differences were identified between the baseline and simple conditions, and between the baseline and complex conditions, which indicates the design of the sound-detection secondary task used in dual-task Experiment 1 was reasonable and sensitive. Furthermore, participants obtained significantly higher accuracy rates on the secondary task when carrying out the hypothesized simple primary writing task, although no evidence of reaction time was detected to support the intended manipulation of task complexity.

DUAL-TASK EXPERIMENT 2

Table 4 summarizes the descriptive statistics for secondary-task reaction time and accuracy rates under the baseline, simple, and complex task conditions, with one participant's data excluded as an outlier.

Prior to conducting one-way repeated-measures ANOVA, the assumptions of normality and sphericity were checked and found to be satisfied. The ANOVA results showed that the effect of task conditions (i.e., baseline, simple, and complex task conditions) on participants' reaction time of the secondary task was significant, $F\,(2, 58) = 67.66$, $p < .001$, $\eta_p^2 = .70$, and with a large effect size. As for accuracy, a significant effect of task

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

TABLE 3. Descriptive statistics for reaction time and accuracy rates across baseline, simple and complex task conditions in dual-task Experiment 1

		Baseline			Simple			Complex		
	N	*M*	*SD*	95% CI	*M*	*SD*	95% CI	*M*	*SD*	95% CI
RT (ms)	30	543.91	123.12	[497.93, 589.88]	848.13	189.61	[777.33, 918.94]	934.77	311.20	[818.56, 1050.97]
Acc (%)	30	98.11	2.17	[97.31, 98.92]	94.19	4.26	[92.60, 95.78]	91.60	4.62	[89.88, 93.33]

Note: RT = reaction time, ms = milliseconds, Acc = accuracy rates.

conditions on participants' performance was identified, $F(2, 58) = 16.58$, $p < .001$, $\eta_p^2 = .36$, and associated with a large effect size. Pairwise comparisons of estimated marginal means with a Bonferroni adjusted α level were applied to investigate where the statistically significant differences were located. The pairwise comparisons revealed that increases in task complexity of the primary writing task had a significant impact on participants' reaction time in the secondary task: simple task ($M = 496.95$, $SD = 86.50$) versus complex task ($M = 533.37$, $SD = 104.71$), $p = .008$, $d = -.37$, with an effect size that ranged from small to medium. Participants also responded to the sound significantly faster under the baseline condition ($M = 385.14$, $SD = 65.33$), than under either the simple primary task condition, $p < .001$, $d = -1.42$, or the complex condition, $p < .001$, $d = -1.58$; the effect sizes were large. These findings indicate that intended increases in task complexity of the primary writing task statistically increased participants' reaction time in the secondary task. Moreover, the design of this version of the secondary task was reasonable and sensitive in terms of reaction time to distinguish the baseline and experimental conditions.

For accuracy rates, intended increases in task complexity of the primary task, in contrast to our expectation, did not contribute to a significant decline: Accuracy in the simple task ($M = 95.47$, $SD = 2.76$) was similar to that in the complex task ($M = 94.04$, $SD = 3.10$), $p = .103$, $d = .49$. Significant differences were, however, detected between the baseline ($M = 97.88$, $SD = 3.14$) and the simple conditions, $p = .001$, $d = .81$ and between the baseline and the complex conditions, $p < .001$, $d = 1.23$; the effect sizes were large. The results suggest that the designed task complexity of the primary task did not statistically affect accuracy, although the baseline and experimental conditions were successfully differentiated by accuracy.

In sum, comparisons of the baseline-simple and baseline-complex conditions indicated that the combination of the two primary tasks and the auditory secondary task in Experiment 2 constituted a relatively sensitive measure of cognitive demands. The results for reaction time of the secondary tasks offer evidence in support of the presumed task complexity manipulations of the two writing tasks; that is the hypothesized complex task was more complex. Although the results for accuracy rates in this experiment were not supportive of the task complexity, the relatively high accuracy rates support the validity of the reaction time data.

DISCUSSION

SELF-RATINGS

The results found that the hypothesized-to-be complex task was perceived as significantly more difficult and requiring significantly more mental effort than the simple task. These findings are in line with previous research (Révész et al., 2016; Sasayama, 2016). For example, in Révész et al.'s (2016) study, participants rated the hypothesized complex task more difficult, consuming more mental effort when completing the map and decision-making tasks than the simple task. Sasayama (2016) found that among the four tasks, the designed simplest task was rated with the lowest level of task complexity and the most complex task with the highest level of cognitive complexity, measured by both self-ratings of cognitive demands.

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

TABLE 4. Descriptive statistics for reaction time, and accuracy rates across baseline, simple and complex task conditions in dual-task Experiment 2

		Baseline			Simple			Complex		
	N	*M*	*SD*	95% CI	*M*	*SD*	95% CI	*M*	*SD*	95% CI
RT (ms)	30	385.14	65.33	[360.74, 409.53]	496.95	86.50	[464.65, 529.25]	533.37	104.71	[494.27, 572.47]
Acc (%)	30	97.88	3.14	[96.70, 99.05]	95.47	2.76	[94.44, 96.50]	94.04	3.10	[92.88, 95.20]

Note: RT = reaction time, ms = milliseconds, Acc = accuracy rates.

The findings in the present study, however, are inconsistent with aspects of Révész et al.'s (2016) study, in which no significant differences were reported between the participants' ratings of task difficulty on the narrative tasks. These are interesting findings, as the difference in outcomes may be due to different task types used in our study and in that of Révész et al.'s (2016). Révész et al. (2016) posited that participants might be more familiar with the narrative task than the map and decision-making tasks. In contrast, the argumentative tasks used in the present study, compared to the narrative, appear to require increased cognitive demands (Andrews, 1995; Freedman & Pringle, 1984), more critical thinking and articulation of ideas, and greater challenges in the use of linguistic elements (Hirose, 2003). Learners may perceive the argumentative tasks as more difficult and consuming more mental effort than the narrative tasks, which may lead to a larger gap of cognitive demands between the simple and complex tasks.

The effect size of task difficulty (d = .52) was larger than that of mental effort (d = .42), although the results of these two measures revealed similar trends across the two writing tasks. This finding suggests that the two variables, task difficulty and mental effort, are related but not identical (Brünken et al., 2010; Révész et al., 2016). Task difficulty is based on learners' perceptions of the cognitive demands of the tasks, determined by the abilities and affective responses that learners bring to the task (Robinson, 2005, 2007a, 2011a), while mental effort is related to the perceived effort that learners used. The larger effect size of task difficulty than mental effort might suggest that, whereas participants perceived the presumed complex task as more difficult than the task designed to be simple, they, however, exerted only a little more effort in the completion of the complex task than the simple one.

The effect sizes for self-ratings in our study were smaller than those in Sasayama's (2016) study (d = .98 for task difficulty and d = 1.00 for mental effort). A possible reason is that Sasayama's results of self-ratings contained extra loads imposed by dual-task methodology, in addition to the task-generated cognitive demands, given participants under the dual-task condition were subjected to the self-ratings. The results in our study represented the cognitive demands yielded by the tasks only, as participants reported their perceptions under the single-task condition.

DUAL-TASK EXPERIMENTS

In dual-task Experiment 1, the results showed that participants had significantly higher accuracy rates on the secondary task when carrying out the simple primary writing task versus the complex one. This finding aligns with previous studies (Révész et al., 2014, 2016), and partially echoes the principle underlying the dual-task paradigm that when the level of cognitive demands imposed by the primary task is increased, less accurate and more reaction time are expected on the secondary task. The results for secondary-task accuracy, therefore, suggest that the hypothesized complex version of the primary writing task in the present study imposed greater cognitive demands than the simple task.

No significant differences were identified for secondary-task reaction time, however, when participants were under the simple primary task condition, as compared to the complex task condition. This finding is not surprising because the same trends were observed in previous research (Révész et al., 2014, 2016; Sasayama, 2016). This is probably because the design of the secondary task in the present study is relatively similar to those used in previous studies (Révész et al., 2014, 2016; Sasayama, 2016), although

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

the modality of the two secondary tasks is different (i.e., one is an auditory task and the others are visual tasks). Specifically, the secondary task in the present study required participants to detect the change of two sounds. The other three studies (Révész et al., 2014, 2016; Sasayama, 2016) directed students to notice the change of two colors and react to only one. An explanation for the non-salient reaction time, as noted by Révész et al. (2016), may be that the secondary-task reaction time is less sensitive than the accuracy rate to the cognitive demands generated by the primary task, when the dual-task condition is operationalized as a particular type of color- or sound-detection task.

Another possible explanation for the lack of simultaneous secondary-task effects for reaction time and accuracy might be the potential competition for attentional resources between reaction time and accuracy. In a dual-task experiment, the primary task is normally given priority and, in this way, the limited residual mental resources that are not being used by the primary task could be allocated to the secondary task (Marcus et al., 1996). For the present and the three aforementioned studies, students had to quickly distinguish whether the sound/color is correct and simultaneously react to the right sound or color when completing the secondary task with limited attentional resources. Completing the secondary task required participants to shift attention from determining whether the stimulus is right or wrong to producing a response to the stimulus. During this process, students may not equally allocate attention to noticing the correct stimulus (accuracy) and responding to the stimulus (reaction time). According to Wickens' (2007) Selective Attention Model, participants' effort of retrieval in the correct sound stimulus might be traded off against the effort of responding to the sound. This suggests that the more effort on reaction time, the fewer differences in reaction time would be found between the simple and complex primary tasks, and accuracy would be curtailed. Given the increased cognitive demands imposed by the primary task, the mental resources for the secondary task would be more limited, and, consequently, the competition might be more intense between the secondary-task reaction time and accuracy. If more attention is directed to reacting to the sound, less attention is allocated to achieving accuracy, and vice versa. Thus, when students complete the complex primary task, two possible result patterns might be found. One is that participants might slightly increase secondary-task reaction time (sometimes reaction time might not be detectable) but statistically decrease accuracy rates, as found in Experiment 1 of our study. Another possible pattern might be, inversely, significantly more reaction time but less detectable changes in accuracy, which was found in Experiment 2 (discussed subsequently). This is probably why when participants completed the designed complex primary writing task in Experiment 1, the supposed-to-be longer reaction time was not found to be statistically increased, but the accuracy rate declined significantly in the secondary task.

With dual-task Experiment 2, the reaction time for the secondary task offered evidence in support of the presumed manipulation of two primary writing tasks. Increases in the task complexity of the primary writing task statistically increased participants' reaction time in the secondary task. No significant decrease, however, was identified for secondary-task accuracy rate when students completed the hypothesized complex primary writing task versus the simple one. Although patterns for reaction time and accuracy rate, found in Experiment 2, were seemingly opposite to those in Experiment 1, the difference might be due to the designs of the secondary task, which is consistent with the speculation in Lee (2019). Different secondary tasks were connected to varying

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

information processing steps, making it possible to identify on which of the processes the cognitive demand is imposed (Baddeley, 1986; Brünken et al., 2003). The secondary task in Experiment 1 was a sound-detection task, in which participants were asked to react fast and accurately to a specific sound and ignore another sound. Participants in Experiment 2 were instructed to respond to only one sound, although the stimulus appeared more frequently in comparison with Experiment 1. The designs of the secondary task led to different information processing steps and directed participants' attention to different aspects, reaction time, or accuracy. As previously discussed, students appeared to focus on reaction time more than accuracy in Experiment 1, leading to significant differences in the secondary-task accuracy rate. A similar example is the color-detection task in previous studies (Révész et al., 2014, 2016). Competition between accuracy and reaction time in the secondary task would suggest that participants directed more attention to accuracy than reaction time when the secondary task is operationalized as a tone-response task in Experiment 2 instead of the stimulus-detection task.

From the comparison between the two dual-task experiments, it seems that the sound-detection secondary task in Experiment 1 had a stronger effect than the sound-response version in Experiment 2 on accuracy rates. The effect size for accuracy in Experiment 1 (d = .58) is larger than that found in Experiment 2 (d = .49), indicating a more detectable difference in accuracy elicited by the sound-detection secondary task. For reaction time, despite the statistically non-significant differences found in Experiment 1, the effect sizes in the two experiments were quite similar (d = .31 in Experiment 1, d = .37 in Experiment 2), suggesting a similar secondary-task effect on reaction time in the two experiments.

The pronounced secondary-task effects in Experiment 1 likely stem from its creating "the required level of interference" (Révész et al., 2016, p. 731). Compared to the sound-response secondary task (responding to the sound only) in Experiment 2, the sound-detection version required participants to both notice and respond to the change of two sounds. When the "required level of interference" has been created in the secondary tasks, competition for attentional resources between the primary and secondary tasks would be greater (Révész et al., 2016), leading to a pronounced secondary-task effect with detectable differences in accuracy or reaction time. In contrast, the sound-response version had a lower level of interference, and therefore, is perhaps not sensitive enough to detect the fine difference in accuracy between the two primary tasks. Reduced competition for attentional resources between the primary and secondary tasks led to participants' rapid responses to the sound (i.e., 848.13 ms for the simple and 934.77 ms for the complex task in Experiment 1, while 496.95 ms for the simple and 533.37 ms for the complex task in Experiment 2). Of note, the effect size for reaction time elicited by the sound-response version is larger than that found for the same measure used in Sasayama (2016; d = .34) and Révész et al. (2014; d = .15 for ESL learners), although the secondary-task effects yielded in Experiment 2 were not as pronounced as those elicited in Experiment 1.

To sum up, the results obtained from the dual-task experiments, that is, accuracy rates in Experiment 1 and reaction time in Experiment 2, closely align with the findings of the self-ratings on task difficulty and mental effort. These results suggest that these indicators measure the same construct, cognitive demands, but from different perspectives (Brünken et al., 2010), which highlights the necessity of triangulation. The findings obtained from the three data sources appear to confirm that the manipulations of task complexity in this

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

research were successful: The complex task version, as intended, yielded greater cognitive demands than the simple version.

CONCLUSION

This study was set up to validate the intended task complexity manipulated within two argumentative writing tasks by varying numbers of argument elements and reasoning demands. The results for the self-ratings on task difficulty and mental effort, and dual-task experiments indicate that the hypothesized complex task promoted more cognitive demands than the task designed to be simple.

The current study is the first that attempts to assess the manipulations of presumed task complexity in L2 writing by using a dual-task method to objectively measure cognitive demands, although some researchers (e.g., Révész et al., 2016; Sasayama, 2016) have pioneered such measurement in the oral tasks in L2 task-based research. Dual-task techniques, as well as self-ratings, are proven to be reliable methods for the validation of the intended task complexity manipulations in writing. Triangulation of the data from different sources is critical to ensuring that the findings for the manipulations of task complexity are accurate and reliable.

Various cognitive-demand measures have their own advantages and disadvantages. Self-ratings might be one of the easiest and most appropriate measures used in L2 classrooms. It is easy for teachers to use the *Task Difficulty and Mental Effort Questionnaire* to understand whether the task complexity manipulations worked as intended. Accordingly, teachers can adjust the cognitive demands of writing tasks to optimize their task-grading and sequencing decisions in L2 teaching pedagogy.

The dual-task method can directly and objectively measure cognitive demands. The dual-task Experiment 1 elicited the largest effect size ($d = .58$ on accuracy) between the presumed simple and complex tasks across all the measures used in this study. The sound-detection design of the secondary task in Experiment 1 may have been sufficiently robust to capture different secondary-task accuracy. This measure could be employed by curriculum designers as well as test assessors to objectively validate cognitive complexity of the L2 writing tasks in the field of curriculum design and writing assessment.

In contrast, the sound-response version might be more suitable for novice L2 learners, given its stronger effect size for reaction time than those found in previous studies (Révész et al., 2014; Sasayama, 2016). The novice learners' attentional limits may be easily reached during L2 writing, and the lower level of interference of the sound-response secondary task may provide as little disturbance as possible to the primary task during the dual-task experiment. This suggestion, however, should be interpreted with caution and further investigated, as participants' proficiency level was not taken into consideration in our study.

Differences in findings of the dual-task Experiment 1 and Experiment 2 in our study highlight the importance of the choice of a secondary task in using dual-task methods, as different versions of secondary tasks may lead to participants' differential attentional allocation and distinct secondary-task effects. Various versions of secondary tasks need to be implemented in the field of SLA to gain deeper insights into selecting appropriate secondary tasks for validating intended task complexity manipulations, especially in L2 writing, in future research.

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

It needs to be pointed out that our study validated the manipulation of intended task complexity based on the data from Chinese EFL learners. No baseline data were collected from native speakers of English, who have relatively few language barriers when writing compositions in English. This means that future studies can extend our study by tracking responses of native speakers to the tasks and use the two sources of the data (from L1 speakers and L2 speakers) to triangulate the manipulation of the task complexity.

Our study focused on only the argumentative genre, which limits the generalizability of findings to other task types, given task type may moderate the effects of task complexity on cognitive demands (Lee, 2019). Future research would benefit from trialing the methods on various task types and multiple complexity levels of each task type.

Furthermore, replication of this research using self-ratings and, in particular, dual-task methods is worthwhile. For future research, the inclusion of open-ended questionnaires or interviews after the dual-task experiments may help researchers record what each participant was doing in the process of writing while carrying out the secondary task. In addition to self-ratings and dual-task methods, a combination of eye-tracking and keystroke (Révész et al., 2019) may be a way to access the information on cognitive demands with a consideration of pauses during the writing process. Other techniques, such as time estimation (Lee, 2020; Sasayama, 2016), time-on-task (Lee, 2020), stimulated recall (Kim et al., 2015), heart rate test, and neuroimaging techniques are also warranted for exploration in future studies to assess the cognitive demands in L2 writing.

NOTES

[1]Participants were instructed to close the E-prime program when they finished the primary tasks in advance to make sure no time left for only responding to the secondary task. Their completion time and secondary-task data would be automatically saved by E-prime.

[2]Participants' (primary) writing task performance in the dual-task and single-task groups were analyzed to ensure that they could complete the writing tasks under each condition. These data were, however, not intended to answer our research questions, and therefore, were not discussed in our study. Participants' writing performance influenced by task complexity will be addressed in another study.

[3]To ensure that Cohen's benchmarks do not underrate the effects in the L2 field (Plonsky & Oswald, 2014), we compared the effect sizes obtained to those on similar relationships in previous L2 research.

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit http://doi.org/10.1017/S0272263121000310.

REFERENCES

Abrams, Z. I. (2019). The effects of integrated writing on linguistic complexity in L2 writing and task-complexity. *System*, *81*, 110–121. https://doi.org/10.1016/j.system.2019.01.009

Andrews, R. (1995). *Teaching and learning argument.* Cassel.

Baddeley, A. D. (1986). *Working memory*. Oxford University Press.

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. https://doi.org/10.1017/S0272263113000429

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

Baralt, M., Gilabert, R., & Robinson, P. (2014). An introduction to theory and research in task sequencing and instructed second language learning. In M. Baralt, R. Gilabert, & P. Robinson (Eds.), *Task sequencing and instructed second language learning* (pp. 1–34). Bloomsbury Academic.

Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, *38*, 53–61. https://doi.org/10.1207/S15326985EP3801

Brünken, R., Seufert, T., & Paas, F. (2010). Measuring cognitive load. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 181–202). Cambridge University Press.

Cho, H. (2015). Effects of task complexity on English argumentative writing. *English Teaching*, *70*, 107–131. https://doi.org/10.15858/engtea.70.2.201506.107

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

Donate, Á. (2018). *Cognitive task complexity, foreign language anxiety and L2 performance in Spanish: A task-based language teaching perspective* [*Unpublished doctoral dissertation*]. Georgetown University, Washington, DC.

Freedman, A., & Pringle, I. (1984). Why students can't write arguments. *English in Education*, *18*, 73–84. https://doi.org/10.1111/j.1754-8845.1984.tb00668.x

Hirose, K. (2003). Comparing L1 and L2 organizational patterns in the argumentative writing of Japanese EFL students. *Journal of Second Language Writing*, *12*, 181–209. https://doi.org/10.1016/S1060-3743(03)00015-8

Housen, A., & Simoens, H. (2016). Introduction: Cognitive perspectives on difficulty and complexity in L2 acquisition. *Studies in Second Language Acquisition*, *38*, 163–175. https://doi.org/10.1017/S0272263116000176

Ishikawa, T. (2007). The effect of manipulating task complexity along the [+here-and-now] dimension on L2 written narrative discourse. In M. D. P. G. Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 136–156). Multilingual Matters.

Kang, S., & Lee, J. H. (2019). Are two heads always better than one? The effects of collaborative planning on L2 writing in relation to task complexity. *Journal of Second Language Writing*, *45*, 61–72. https://doi.org/10.1016/j.jslw.2019.08.001

Kim, Y. J., Payant, C., & Pearson, P. (2015). The intersection of task-based interaction, task complexity, and working memory: L2 question development through recasts in a laboratory settings. *Studies in Second Language Acquisition*, *37*, 549–581. https://doi.org/10.1017/S0272263114000618

Kuiken, F., Mos, M., & Vedder, I. (2005). Cognitive task complexity and second language writing performance. *EUROSLA Yearbook*, *5*, 195–222. https://doi.org/10.1075/eurosla.5.10kui

Kuiken, F., & Vedder, I. (2007). Cognitive task complexity and linguistic performance in French L2 writing. In M. D. P. G. Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 117–135). Multilingual Matters.

Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing*, *17*, 48–60. https://doi.org/10.1016/j.jslw.2007.08.003

Kuiken, F., & Vedder, I. (2012). Syntactic complexity, lexical variation and accuracy as a function of task complexity and proficiency level in L2 writing and speaking. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 143–170). John Benjamins Publishing Company.

Kuiken, F., & Vedder, I. (2019). Investigating teachers' perceptions of syntactic complexity in L2 academic writing. *Instructed Second Language Acquisition*, *3*, 228–248. https://doi.org/10.1558/isla.37977

Lee, J. (2019). Task complexity, cognitive load, and L1 speech. *Applied Linguistics*, *40*, 506–539. https://doi.org/10.1093/applin/amx054

Lee, J. (2020). Task closure and task complexity effects on L2 written performance. *Journal of Second Language Writing*, *50*, 1–13. https://doi.org/10.1016/j.jslw.2020.100777

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

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology*, *88*, 49–63. https://doi.org/10.1037//0022-0663.88.1.49

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

Ong, J., & Zhang, L. J. (2010). Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing*, *19*, 218–233. https://doi.org/10.1016/j.jslw.2010.10.003

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

Ong, J., & Zhang, L. J. (2013). Effects of the manipulation of cognitive processes on EFL writers' text quality. *TESOL Quarterly*, *47*, 375–398. https://doi.org/10.1002/tesq.55

Paas, F., & Van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, *6*, 351–371. https://doi.org/10.1007/bf02213420

Paas, F., van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, *79*, 419–430. https://doi.org/10.2466/pms.1994.79.1.419

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912. https://doi.org/10.1111/lang.12079

Rahimi, M. (2019). Effects of increasing the degree of reasoning and the number of elements on L2 argumentative writing. *Language Teaching Research*, *23*, 633–654. https://doi.org/10.1177/1362168818761465

Rahimi, M., & Zhang, L. J. (2018). Effects of task complexity and planning conditions on L2 argumentative writing production. *Discourse Processes*, *55*, 726–742. https://doi.org/10.1080/0163853X.2017.1336042

Rahimi, M., & Zhang, L. J. (2019). Writing task complexity, students' motivational beliefs, anxiety and their writing production in English as a second language. *Reading and Writing*, *32*, 761–786. https://doi.org/10.1080/0163853X.2017.1336042

Reed, W. M., Burton, J. K., & Kelly, P. P. (1985). The effects of writing ability and mode of discourse on cognitive capacity engagement. *Research in the Teaching of English*, *19*, 283–297. https://doi.org/10.2307/40171051

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. https://doi.org/10.1093/applin/amt039

Révész, A., Kourtali, N. E., & Mazgutova, D. (2017). Effects of task complexity on L2 writing behaviors and linguistic complexity. *Language Learning*, *67*, 208–241. https://doi.org/10.1111/lang.12205

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. https://doi.org/10.1017/S0272263115000339

Révész, A., Michel, M., & Lee, M. (2019). Exploring second language writers' pausing and revision behaviors. *Studies in Second Language Acquisition*, *41*, 605–631. https://doi.org/10.1017/s027226311900024x

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. https://doi.org/10.1111/lang.12061

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, *22*, 27–57. https://doi.org/10.1093/applin/22.1.27

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching*, *43*, 1–32. https://doi.org/10.1515/iral.2005.43.1.1

Robinson, P. (2007a). Criteria for classifying and sequencing pedagogic tasks. In M. D. P. G. Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7–26). Multilingual Matters.

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

Robinson, P. (2011a). Second language task complexity, the cognition hypothesis, language learning, and performance. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 3–38). John Benjamins Publishing Company.

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

Sasayama, S. (2016). Is a "complex" task really complex? Validating the assumption of cognitive task complexity. *Modern Language Journal*, *100*, 231–254. https://doi.org/10.1111/modl.12313

Sasayama, S., & Norris, J. (2019). Unravelling cognitive task complexity: Learning from learners' perspectives on task characteristics and second language performance. In Z. E. Wen, & M. J. Ahmadian (Ed.), *Researching L2 task performance and pedagogy: In honour of Peter Skehan* (pp. 95–132). John Benjamins Publishing Company.

Schoor, C., Bannert, M., & Brünken, R. (2012). Role of dual task design when measuring cognitive load during multimedia learning. *Educational Technology Research and Development*, *60*, 753–768. https://doi.org/10.1007/s11423-012-9251-8

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics*, *17*, 38–62. https://doi.org/10.1093/applin/17.1.38

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

Skehan, P. (2013). Tasks and language performance assessment. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Research pedagogic tasks: Second language learning, teaching, and testing* (pp. 167–185). Routledge.

Skehan, P. (2018). *Second language task-based performance: Theory, research, assessment*. Routledge.

Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183–205). Cambridge University Press.

Wang, L., & Zhang, L. J. (2019). Peter Skehan's influence in research on task difficulty: A bibliometric analysis using CiteSpace. In Z. E. Wen & M. J. Ahmadian (Eds.), *Researching L2 task performance and pedagogy: In honour of Peter Skehan* (pp. 183–196). John Benjamins Publishing Company.

Wickens, C. (2007). Attention to the second language. *International Review of Applied Linguistics in Language Teaching*, *45*, 177–191.

Zhang, L. J., Wang, L., & Wu, H. (2017). A meta-analysis of linguistic complexity research (1990–2015) conducted within cognitive linguistics frameworks. *Fudan Forum on Foreign Languages and Literature*, *10*, 58–69.

https://doi.org/10.1017/S0272263121000310 Published online by Cambridge University Press