

RESEARCH ARTICLE

Working memory, L2 proficiency, and task complexity: Independent and interactive effects on L2 written performance

Rosa Maria Manchón¹, Sophie McBride¹ , María Dolores Mellado Martínez¹ and Olena Vasylets²

¹University of Murcia, Spain; ²University of Barcelona, Spain

Corresponding author: Sophie McBride; Email: sophie.mcbride@um.es

(Received 07 August 2022; Revised 09 February 2023; Accepted 06 March 2023)

Abstract

This study examined the independent effects of working memory (WM) and the interactive effects of WM/L2 proficiency and WM/task complexity on L2 written performance. The study followed a within-between-participant factorial design, with two levels of task complexity as the within-participant variable and L2 proficiency and WM as between-participants variables. The outcome measure was L2 writing performance as measured by CAF indices. Two groups of undergraduate students from a degree in English studies were invited to complete the simple and complex version of the “Fire-Chief” task. Task complexity was operationalized in terms of reasoning demands, and tasks were counterbalanced to avoid unwanted order effects. Participants also completed the Oxford Placement Test and a working memory test (*n*-back). Regarding independent effects, results show that WM did not have an effect on L2 writing performance. In contrast, L2 proficiency was the variable most connected to various dimensions of the text produced. As for interactive effects, no significant interaction between WM, proficiency, or task complexity was found. In contrast, L2 proficiency emerged as the sole significant predictor of L2 writing performance at both levels of task complexity.

Introduction

The purpose of the study reported in this contribution to the special issue was to add to the expanding body of scholarly work on the role of individual differences (IDs) in the modality of writing by shedding further light on (a) how working memory (WM) is implicated in written language use and (b) whether learner language proficiency and task complexity moderate any potential WM effects.

The motivation for investigating this triple interaction among working memory, L2 proficiency, and task complexity is grounded in recent theoretical proposals and empirical findings suggesting that WM effects may depend on other moderating factors (e.g., Baddeley, 2015; Olive, 2011), especially the level of L2 proficiency in the case of

WM effects on L2 use. For instance, Serafini and Sanz (2016) studied morphosyntactic development in Spanish L2 over the course of a semester and found that, although some components of WM did have an effect on performance in the case of the lower proficiency participants in their study, WM effects diminished as proficiency increased. Such predicted and attested interaction of WM and other variables can be anticipated to be especially relevant in the case of L2 production on account of the cognitively demanding nature of composing, especially in connection with the orchestration of knowledge resources and skills that are implicated in text production. Thus, performing writing (in both time-constrained and time-expanded, individual and collaborative conditions) entails the availability of and (ideally automatic) access to required L2 knowledge, knowledge of genre conventions and rhetorical requirements (Schoonen *et al.*, 2011), and domain knowledge relevant to the task at hand. Writing also presupposes multiple cognitive skills in order to successfully orchestrate (and shift between) the higher order processes involved in writing (essentially planning, linguistic encoding, revision, and monitoring), which entails *inter alia* decision making on the part of the writer as to the allocation of attentional resources throughout the entire process of composing. The implication of WM is thus crucial in managing the cognitively demanding and problem-solving nature of composing, as WM is “the place where writing processes are activated and coordinated and where the writer’s representation of the text is constructed. [It is] the cognitive space where operations of the writing process take place” (Olive, 2011, p. 485). WM functions are essential in writing as, on one hand, the storage function facilitates “temporary stores for transient information created during composing (Olive, 2022, p. 504) and, on the other hand, the processing component is heavily involved in the “coordination and switching among the writing processes, construction of the different representations necessary to create written discourse, and particularly construction of the writer’s multidimensional representation of the text (Olive, 2022, p. 505). This explains the central role attributed to WM in models of writing (Hayes, 1996; Kellogg, 1996, 2001; McCutchen, 1996), as more fully discussed below.

As language is heavily involved in the unavoidable writing process of formulation—or linguistic encoding—in the case of L2 writing, the greater availability of/more automatic access to required L2 knowledge the writer has, the more attentional resources and processing capacity s/he will have to devote to processes other than linguistic encoding (Weigle, 2005). Thus the relevance of looking into the potential interaction between WM and L2 proficiency in L2 written production. Yet, research has looked into the interaction of WM with variables other than L2 proficiency in the domain of L2 writing. This is justified on account of previous work on the independent and interactive effects of task-related and writer-related factors on the above-mentioned orchestration and implementation of writing processes, on one hand, and on the characteristics of the texts produced, on the other. For instance, there is ample empirical evidence showing that task characteristics (Barkaoui, 2016; Michel *et al.*, 2019, 2020; van Weijen, 2009) and individual factors such as proficiency (Barkaoui, 2019; Gánem-Gutiérrez & Gilmore, 2018; Manchón & Roca de Larios, 2007; Roca de Larios *et al.*, 2008; Tillema, 2012) or WM capacity (Michel *et al.*, 2019; Révész *et al.*, 2017) do have an effect on the implementation and temporal distribution of writing processes.

From the point of view of effects of writer-internal and writer-external variables on the texts produced (the focus of the research reported in this paper), the choice of the predictor variables in our study (WM, L2 proficiency, and task complexity) was further motivated by the consideration of previous second language acquisition (SLA) empirical evidence on (a) the role of task complexity in written performance (e.g., Vasylets

et al., 2017; Zalbidea, 2017), (b) the proficiency dependency of L2 users' perception of task complexity (Sasayama, 2016) and of task complexity effects (e.g., Ishikawa, 2007; Kuiken et al., 2005; Kuiken & Vedder, 2008), (c) the limited research and (at times contradictory) available empirical findings on the interaction between WM effects and task complexity (e.g., Kormos & Trebits, 2011; Zalbidea, 2017), and on the L2 proficiency dependency of WM effects (Kormos & Sáfár, 2008; Lu, 2015; Vasylets & Marín, 2021) in the writing domain. The empirical evidence in these various strands is synthesized in the background to the study that follows.

Background

Working memory and writing

Models of L1 writing (e.g., Hayes, 2012; Kellogg, 1996, 2001) view WM as a central cognitive resource for composing. The implication of WM components is predicated on, first, the cognitive demanding nature of composing and, second, the consideration that writing entails the different functions of WM—that is, the storing and processing components. Based on Baddeley's (1986) multicomponential model of WM, Kellogg's (1996) L1 writing model (which has informed most studies on WM in L2 writing) establishes relations between WM components and writing processes. Thus, the central executive is purported to be implicated in all higher level writing processes, which in his model are formulation (including planning and linguistic encoding), execution, and monitoring (including revision and editing). In contrast, the visual-spatial sketchpad is related to just planning, whereas the phonological loop is purported to be linked to the processes of translation and revision. Theoretical predictions on the role of WM on L1 writing have been confirmed empirically, as WM has been shown to play an essential role in older and younger writers' L1 writing performance (see Olive, 2011, 2022, for reviews. See also Kormos's and Li's contributions to this special issue).

In her pioneering account of IDs and L2 writing, Kormos (2012) convincingly argued the implication of WM capacity in all stages of composing. She did so on account of (a) the attested cognitive demanding nature of writing in general and (b) the extra additional demands on cognitive resources that writing in an additional language (L2) may impose as a result of lack of (automatic access to) relevant L2 knowledge needed to convey one's intended meaning. In fact, research on L2 writing processes has provided ample empirical evidence of the more labor intensive nature of linguistic encoding in L2 writing compared with writing in one's native language. For instance, in a synthesis of their process-oriented studies intended to shed light on L2 writers' problem-solving behavior while composing in their L1 and L2, Manchón et al. (2009) confirmed the anticipated quantitative and qualitative differences in L1 and L2 writing-processing activity resulting from differences and accessibility of L1 and L2 knowledge. They found that, although across languages and proficiency levels most composing time was devoted to linguistic encoding, this process of transforming ideas into language was significantly more fluent (as opposed to involving different degrees of problem solving) in L1 than in L2, thus providing support for the prediction that writing in an additional language imposes a heavier burden on writers (see also Roca de Larios et al., 2008).

Kormos (2012) thus hypothesized WM effects in the implementation of writing processes and in the characteristics of the texts produced. These assumptions and predictions have been partially confirmed in a body of empirical work, the findings of which, globally considered, are nevertheless unclear and at times contradictory (see

Kormos's and Li's detailed reviews in this special issue). More precisely, whereas some previous studies have shown that there is a positive relationship between working memory capacity and L2 writing quality (e.g., Adams & Guillot, 2008; Baoshu & Chuanbi, 2015; Kormos & Sáfár, 2008; Mavrou, 2020; Mujtaba *et al.*, 2021; Peng *et al.*, 2022; Révész *et al.*, 2017; Vasylets & Marín, 2021; Zalbidea, 2017), others have found mixed results (Bergsleithner, 2010; Michel *et al.*, 2019; Zabihi, 2018) or practically null effects (Cho, 2018; Lu, 2015).

The contradictory nature of available empirical evidence is even more evident when we consider the linguistic dimensions of texts found to be affected by WM, as reported in a body of studies looking at WM effects on the texts produced by adolescent and young adult L2 users with diverse L1 backgrounds learning English or Spanish as an L2. In an early WM study by Adams and Guillot (2008) with French/English bilinguals, the researchers found an effect of the phonological component of WM on the participants' L2 texts, especially in the area of spelling. Participants' phonological short-term memory capacity was assessed by a listening-span task. L2 writing performance was assessed by holistic ratings, but no measures of fluency were employed.

WM effects were also reported by Bergsleithner (2010), who found a positive effect of WM on the accuracy and subordination of the English L2 texts written by L1 Brazilian participants, two areas also found to be affected by WM in the texts written by the Spanish L2 learners in Mavrou's (2020) study. Bergsleithner (2010) used an operation–word span test to measure WM, whereas Mavrou (2020) employed five WM tasks, including a backward Corsi block-tapping task, an operation span task, a running-memory-span task, a number–letter task, and an emotional Stroop task. In Mavrou's (2020) study, fluency was operationalized in terms of the total number of tokens, T-units, and clauses (Wolfe-Quintero *et al.*, 1998), whereas in the study by Bergsleithner (2010) no measures of fluency were employed. WM effects on the accuracy dimension of texts were also reported by Baoshu and Luo (2012), Zalbidea (2017), and Mujtaba *et al.* (2021). In these three studies, WM was assessed by means of the operation span task. In Baoshu and Luo's (2012) study, fluency was measured by words per minute, whereas the studies by Zalbidea (2017) and Mujtaba *et al.* (2021) did not measure fluency.

In contrast to these findings, in a study with Persian learners of L2 English, Zabihi (2018) reported positive WM effects on fluency (operationalized as the number of words per T-unit) and subordination, but not in accuracy. WM was also measured by an operation span task. In short, rather mixed findings in the domains of accuracy, fluency, and syntactic complexity have been reported.

Regarding the dimension of lexis, research (with adult L2 users) points to a link between some components of working memory and lexical sophistication (e.g., Vasylets & Marín, 2021) but not with lexical diversity (Mavrou, 2020; Vasylets & Marín, 2021). Yet, the picture appears to be even more complex: Zalbidea's study (2017) of WM effects in speaking and writing as a function of task complexity with Spanish L2 learners found that written production proved to be more lexically complex and more accurate overall than oral production, although only the dimension of accuracy, not that of lexical complexity, was related to WM capacity.

In addition to these mixed findings on the dimensions of performance that are/are not affected by L2 writers' WM, Vasylets and Marín (2021) have convincingly pointed to a methodological issue (in part a limitation) in some previous research. Their concern relates to the way in which the outcome variable of text characteristics has been operationally defined and measured. They noted that some previous studies had somewhat failed to account for the multidimensional nature of crucial linguistic

dimensions of performance, especially regarding complexity. For instance, they argued, syntactic complexity had been studied mainly in terms of subordination (as evidenced in the above synthesis of findings), thus leaving out important syntactic subdimensions such as coordination or nominal complexity.

In sum, the existing empirical work does point to a role for working memory in explaining the linguistic features of L2 written texts, although mixed findings exist on the specific dimensions of writing found to be affected by WM. Given the inconclusive nature of past research, and given also the suggested relevance of expanding the spectrum of L2 written performance dimensions to be investigated in order to gain more nuanced understandings of potential WM effects on production, the first objective of our study was to investigate potential WM effects on a range of complexity, accuracy, and fluency (CAF) dimensions.

To shed further light on the intricacies of the link between WM and writing, and on account of the above-mentioned potential moderating role of additional variables on WM effects on written output, some studies have additionally inspected WM effects as a function of either learner-related variables (essentially, proficiency) or task-related variables (focusing primarily on task complexity). This research, to be reviewed next, provides the motivation for the second and third global objectives of our study.

Working memory and L2 writing: Moderating effects of proficiency

As in the case of global WM effects, the research investigating proficiency-related WM effects have brought about contrasting results. Kormos and Sáfár's (2008) study of 121 Hungarian (beginner and preintermediate) secondary school learners of EFL showed a positive, moderate correlation between phonological short-term memory and L2 writing (operationalized as a holistic rating measure), although the effect was observed only for the preintermediate learners in the study (a negative, nonsignificant correlation was found with beginners' writing performance), and no significant effects of WM on L2 writing were reported for the lower, beginner learners. In contrast, in Lu's (2015) study of 136 Chinese university students, no relationship between WM and L2 writing was detected for either proficiency group. Yet, there are important methodological differences between these two studies worth pointing out. First, they differed in how they measured the predictor variables of proficiency and WM: proficiency was measured by a standardized test in Kormos and Sáfár's (2008) study (the Cambridge First Certificate Exam), whereas Lu's (2015) participants were assigned to a high or low proficiency group on the basis of their scores on a receptive and productive vocabulary test. Similarly, WM was assessed by an operation-span task (involving addition, subtraction, multiplication, and division) in Lu's (2015) research, whereas Kormos and Sáfár (2008) used a nonword span test to measure their participants' phonological short-term memory capacity and a backward digit span test to assess working memory capacity (although the latter was only administered to the beginner participants in the study). Second, the studies also varied in terms of number and genre of texts written by the participants: The beginning and preintermediate secondary school writers in Kormos and Sáfár's (2008) study completed three writing tasks representing different genres, whereas Lu's (2015) participants wrote just one argumentative essay in L2 English under a timed condition. Third, there were also differences in the measurement of the outcome variable: Lu (2015) used an analytic rubric for the dimension of language use (the study also looked into effects on content and organization), whereas Kormos and Sáfár (2008) used holistic ratings, thus using a measure of overall writing

quality. Therefore, the divergent findings can be in part explained by methodological differences.

More recently, Vasylets and Marín (2021) provided evidence of the influence of L2 proficiency on a range of outcome measures in terms of holistic ratings and quantitative CAF measures, the latter including linguistic and propositional indices of complexity. The participants were 59 native Spanish/Catalan learners of L3 English with degrees of L3 proficiency ranging from B1 to C2 according to the Common European Framework of Reference for Languages (Council of Europe, Council for Cultural Cooperation, Education Committee, & Modern Languages Division, 2001). WM capacity was measured via a complex verbal-span task (Gilbert & Muñoz, 2010) for L1 Spanish/Catalan speakers—a Spanish/Catalan version of Unsworth *et al.*'s (2005) operation span task and L2 proficiency via the pen-and-paper version of the Oxford Quick Placement Test, UCLES, 2001. The study provided telling evidence of the degree of complexity involved in ascertaining L2 proficiency effects, clearly pointing to interactions between WM and proficiency depending on which performance measures are considered. Thus, they found a positive link between WM capacity and the dimensions of accuracy and lexical sophistication, although this interaction was moderated by proficiency. The study showed that higher WM capacity was positively related to accuracy but only for low proficiency writers, a finding that they interpreted as suggesting that “writers with higher WMC [working memory capacity] would find themselves better equipped to compensate for gaps in L2 proficiency, successfully resolving various linguistic challenges related to the ability to communicate without errors” (p. 9). Interestingly, the opposite trend was observed in the case of WM effects on lexical sophistication, as WM correlated positively with lexical sophistication only for the participants with higher L2 proficiency in the study. The researchers interpreted this finding by suggesting a link between lexical complexity and higher-order writing processes (especially formulation and monitoring) in which WM is clearly implicated. For instance, in the case of formulation, they speculated with the possibility that greater WM capacity “could facilitate the preparation of a complex conceptual plan, calling for more lexically sophisticated linguistic encoding” (p. 10), a process that requires the use of strategies (for instance, certain lexical searches) that perhaps are only available for use when a certain level of L2 proficiency has been reached and L2 writers have the necessary lexical resources to draw on. Accordingly, they concluded, for their lower proficiency writers, WM did not have an effect on lexical sophistication simply because “their vocabulary was not sophisticated enough for WM to make a meaningful impact” (p. 10).

Given these diverse and at times contradictory findings, further research intended to shed light on the attested complex interaction between WM and proficiency effects on written performance is justified. To advance in this direction, the second objective of our study was to investigate whether any observed WM effects on CAF measures were moderated by proficiency. On account of Vasylets and Marín's (2021) evidence of the differential effect of WM on diverse dimensions of production across proficiency levels, we inspected the interactive effects of WM and L2 proficiency in a wide range of CAF indices.

Working memory and L2 writing: Moderating effects of task complexity

Task complexity refers to the cognitive load of task performance (Sasayama, 2016). In Robinson's (2001) well-known characterization, task complexity is defined as “attentional, memory, reasoning, and other information processing demands imposed by the

structure of the task on the language learner” (p. 29). In the oral domain, a general finding is that task complexity effects on performance are stronger than the effects of L2 proficiency or WM capacity (e.g., Awwad & Tavakoli, 2022; Kormos & Trebits, 2011), although these studies also reported interactions between task complexity and working memory. For instance, Awwad and Tavakoli (2022) found that WM predicted accuracy in more and less complex tasks (operationalized in terms of varying degrees of intentional reasoning) and lexical complexity in the more complex task.

The rationale for a predicted interaction of WM and task complexity in the case of written production rests on the consideration of the cognitively demanding nature of composing referred to above. Thus, the greater the cognitive load of the task, the more involvement of working memory in the orchestration of writing processes, which, in turn, would likely influence the outcome of the process in terms of text characteristics. Accordingly, it has also been argued that WM executive functions might be differentially involved when composing different types of text and that, accordingly, WM research “should focus on a wider variety of texts and writings to examine whether the involvement of working memory varies” (Olive, 2022, p. 517). Along the same lines, McCormick and Sanz (2022) claim that the role of MW is “contingent on task characteristics that challenge learners’ storage and processing capacities” (p. 575).

These predictions have been tested in a handful of studies with younger and older learners (e.g., Michel et al., 2019; Zalbidea, 2017), which found that task complexity factors moderated WM effects on written performance (although the effects were at times small). In Zalbidea’s (2017) study with 32 intermediate learners of Spanish, WM (as measured by an operation span test) effects on the accuracy of L2 written argumentative performance were observed when the cognitive demands of tasks were increased. However, no correlations between WM and lexical and syntactic complexity were found. In Michel et al.’s (2019) study with young writers, a surprising finding was the lack of WM effects on the participants’ performance in most of the writing tasks they were invited to complete, which were four writing task types as part of the TOEFL Junior Comprehensive test battery—namely, an editing task (error correction in a paragraph of a nonacademic and an academic text), an email task (reply to an email), an opinion task (expressing opinion on a topic in 100–150 words), and an integrated listen–write task (writing of a summary paragraph after listening to a teacher talking for approximately 90 s about an academic topic with the help of visual input). The researchers found an effect only for the academic version of the editing task (one in which participants were required to find and correct errors in an academic text) and the integrated listen–write task, although in this case the effect was only found for one proficiency level. The findings on the editing task are in line with Zalbidea’s (2017) WM effects on accuracy, as are also the observed “non-significant, but meaningful” (p. 42) WM effects in the listening-to-write task. The latter effects point to a moderating effect of task complexity on WM effects, as this was considered a “complex task type, as it requires young learners to recall and summarize the content of aural input with support from visual input” (p. 42). In line with some arguments presented in previous sections on the implication of WM in the orchestration of writing processes, the authors speculated with the possibility that the successful execution of the processes involved in their listen–write task may have been assisted by the participants’ WM capacity to coordinate attentional processes. Importantly, the study also found that “learners with high WM functions showed somewhat more consistent performance across tasks than did learners with low WM functions” (p. 43), which could point to another dimension of the association between WM and task characteristics.

Motivated by these mixed findings and the limited research on the topic, the third objective of our study was to examine the link between potential WM effects on the CAF dimensions of the text written by higher and lower proficiency L2 writers and the complexity of the task to be performed. A further motivation for this research objective derives from the consideration of previous SLA work on the interaction between task complexity effects on language use and L2 user's proficiency level, as briefly synthesized next.

Interaction between task complexity and proficiency in L2 writing

An SLA-oriented L2 writing research strand has investigated the interactive effects of task complexity and proficiency in writing. Regarding effects on texts, this research once again shows mixed findings across and within the text dimensions in focus. Concerning accuracy, the general finding is that L2 writers produce more accurate texts when completing more complex tasks. Yet, different results have been obtained when L2 proficiency was included in the analysis, with studies showing no interaction between task complexity and proficiency (e.g., Kuiken *et al.*, 2005) as well as a significant interaction (e.g., Ishikawa, 2007). More consistent findings exist for complexity, especially lexical complexity, as both Ishikawa (2007) and Kuiken and Vedder (2008) reported a proficiency dependency of task-complexity effects on the lexical complexity of L2 production. This attested potential interaction of proficiency and task complexity provides additional motivation for our goal to zoom into the interactive effects of working memory, task complexity, and proficiency.

Research questions

Based on the preceding literature review, it is pertinent to examine potential independent and interactive effects of working memory, L2 proficiency, and task complexity on L2 written performance. Accordingly, our study sought to answer the following research questions:

RQ 1: To what extent does working memory affect L2 written performance, operationalized in terms of complexity, accuracy, and fluency indices?

RQ 2: Do any observed working memory effects on L2 written performance vary as a function of writers' L2 proficiency?

RQ 3: Do any observed working memory effects on L2 written performance vary as a function of the cognitive complexity of the writing task?

Method

Design

The study followed a within-between-participant factorial design, with two levels of task complexity as the within-participant variable and L2 proficiency and WM as between-participants variables. The outcome measure was L2 writing performance as measured by CAF indices.

Participants

The participants were 76 (59 female and 17 male) Spanish undergraduate students majoring in English studies at a Spanish university with different L2 proficiency levels. The participants mean age was 19.8 ($SD = 1.9$, range: 17–25).

Instruments

The writing task

The participants were invited to complete the complex and simple versions (operationalized in terms of reasoning demands) of the “Fire-Chief” task (Gilabert, 2005). This task is a problem-solving, picture-based writing activity in which students are presented with an image of a burning building from which numerous people need to be rescued (see Appendix 1 and Appendix 2 for the complex and simple version, respectively). The design of this task springs from the crisis management simulation in which teams of emergency experts discuss potential crisis situations and provide suggestions of the adequate actions (Gilabert, 2007). Thus, this type of task can be performed in both the oral and written modalities. In terms of the specific operationalization of task complexity, the simple and complex versions of the Fire Chief task represent a visual prompt of a building on fire, various characters (e.g., an old man, a pregnant woman, etc.) trapped in the building, and the available resources (e.g., helicopter, fire truck). The instructions of the task, which are identical in both the simple and complex conditions, ask the participants to explain (a) what actions they would take in order to save as many people as possible from the burning building, (b) in what order they would rescue these people, and (c) why they would take these actions. The major distinction between the simple and the complex versions lies in the amount of resources (the simple version provides a greater amount of rescue resources) and level of the danger (the situation is more critical in the complex task with some vulnerable characters, such as a pregnant woman, exposed to the imminent danger). These differences between the conditions are expected to pose higher levels of cognitive load in the complex task condition. By employing dual-task methodology and self-ratings, Révész et al. (2016) provided empirical evidence for the greater cognitive complexity of the complex version of the Fire Chief task as compared with the simple task condition. Although the results in Révész et al. (2016) were restricted to the performance of the tasks in the oral mode, we consider their results relevant to implement the task as a valid one for the present study. In this respect, Vasylets et al. (2017), who also employed the task in both oral and written modalities, asked their participants to self-assess the cognitive load posed by the task using a 9-point Likert-type scale. The results showed that the ratings of the cognitive load were significantly higher in the complex task condition in the oral and written modalities. These findings are taken as validation for the use of the Fire Chief task in both the oral and written modalities.

Measure of L2 proficiency

To assess the proficiency level of the participants, the classic version of the Oxford Placement Grammar Test (OPT) was administered (Allen, 1992). The grammar section of the OPT consists of 100 questions on grammar knowledge, including fill-in-the-gap exercises and multiple-choice questions, which assess the test taker’s proficiency level according to the Common European Framework of Reference for Languages. Therefore, the results obtained from the test can range from an A1 level

to a C2 level. The participants obtained an average proficiency score of 77.48 ($SD = 9.59$), with the scores ranging from 45 to 95.

Measure of working memory

To assess WM, we employed the *n*-back working memory test (Kane et al., 2007). This test, administered online via <https://www.psytoolkit.org/> (Stoet, 2010, 2017), has been used and validated in previous cognitive research in psychology and neuroscience fields and has been found to be an appropriate instrument for measuring WM (see Conway et al., 2005; Jaeggi et al., 2010; Kane et al., 2007). Similar to other WM tests used in previous research (e.g., Kim et al., 2021), the *n*-back working memory test consists of the provision of a sequence of stimuli (each lasting a few seconds) in the form of letters. Participants are required to decide whether the stimulus they are presented with on the screen is the same letter that they had viewed three trials previously (3-back test). Results for correct answers and errors made are computed by calculating the raw numbers of the correct responses and errors. The participants obtained a mean WM score of 1.03 ($SD = 0.75$; range: 0.30–3.26). The (computerized version of the) *n*-back test was selected over other WM tests because it taps into the maintenance and temporary storage, continuous updating, and processing of information in WM (Gajewski et al., 2018), which represent the functionality of WM relevant for writing, as discussed in earlier sections.

Data collection procedures

Data were collected over the course of five 50-min sessions for both 1st- and 4th-year groups. In the first session (Day 1), the participants were invited to complete the OPT Grammar test (Allen, 1992) in order to confirm L2 proficiency homogeneity within groups and differences across groups. In session two (Day 2), participants were asked to compose their response to the “Fire-Chief” task (Gilbert, 2005) and were divided into two groups, with half of the students completing the simple version task and the remaining half completing the complex version. Each participant received the task prompt and instructions for completion as well as a blank sheet on which to write their texts. Prior to starting writing, participants were asked to read the instructions carefully and to familiarize themselves with the picture so as to get an overall idea of the situation presented in the task. Participants were given 50 min to compose their text, with no specific word limit established. The following 50-min session (Day 3) took place in the computer lab and involved the collection of WM data through the completion of the *n*-back working memory test. The fourth and final session (Day 4) invited students back to complete their second composition, with tasks being counterbalanced between this and the first writing session.

Data analyses

L2 written production: CAF measures

The participants’ L2 written production was analyzed for complexity, accuracy, and fluency measures (see Table 1). To assess L2 writing accuracy, errors (including grammatical, lexical, spelling, and punctuation) were identified and the ratio of errors per 100 words was calculated: total number of errors/total number of words $\times 100$. Two authors analyzed the data, and intercoder reliability for the identification and classification of errors was 96%.

Table 1. Summary of CAF measures used in the study.

Lexical complexity	Syntactic complexity	Accuracy	Fluency
Lexical density (LD)	Mean length of T-unit (MLT)	Ratio of errors x 100	Total number of words
Lexical sophistication (LS2)	Mean length of clause (MLC)		Words per minute
Lexical diversity (UBER index)	Coordinate phrase per clause (CP/C) Dependent clause per clause (DC/C) Complex nominal per clause (CN/C)		

Fluency measures were calculated by computing the total number of words written per minute (total words/total time) and total number of words following previous research (e.g., Wolfe-Quintero et al., 1998). Task composition time was measured by noting down the exact time students commenced and finished writing (within the maximum 50-min allocated writing time). Mean time spent on the complex task amounted to 1,133.6 s ($SD = 393$; range: 300–1,980 s); on the simple task, the participants spent an average of 976.5 s ($SD = 332$; range: 360–1,860 s), about 15 min and 18 min, respectively.

The written texts were analyzed in terms of lexical and syntactic complexity using Synlex software (Lu, 2010). Following Read (2000), lexical complexity was conceptualized as a multidimensional feature consisting of several interrelated components, including lexical density, sophistication, and variation. Lexical density was computed as the ratio of the number of lexical words (as opposed to grammatical words) to the total number of words. In the Lexical Complexity Analyzer, lexical words include nouns, adjectives, lexical adverbs, and verbs (excluding modal verbs, auxiliary verbs, “be,” and “have”); an adverb is considered a lexical adverb if it also appears as an adjective in British National Corpus word list (Leech et al., 2001) or if it consists of an adjectival root and the “ly” suffix. Lexical sophistication was measured by calculating the proportion of relatively advanced/sophisticated words in the learner’s production to the total number of words. Synlex considers a word as sophisticated if it is not among the 2,000 most frequent words in the British national word list. For lexical variation, defined as the range of a learner’s vocabulary as displayed in the language production, UBER index (Dugast, 1979) was obtained in Synlex.

For syntactic complexity, Synlex was used (a) to calculate the mean length of T-units, which was employed as a general measure of complexity; (b) to assess complexity via coordination (coordinate phrases/total number of clauses); and (c) to measure complexity via subordination (dependent clauses/total number of clauses). Regarding nominal complexity, the mean length of clause was calculated to tap into phrasal complexity; also, the ratio of complex nominal structures (complex nominals/total number of clauses) was calculated. Following Lu (2011), complex nominals included (1) nouns plus adjective, possessive, prepositional phrase, adjective clause, participle, or appositive, (2) nominal clauses, and (3) gerunds and infinitives in subject position.

Statistical analyses

All statistical analyses were carried out using the Statistical Package for the Social Sciences (SPSS IBM v28). Given the independent/moderator and dependent variables

in the design of the study, descriptive statistics were computed for working memory and L2 proficiency. We carried out correlations (as a preliminary analysis) as well as regressions. Correlations were performed between the predictor variables of working memory and proficiency, and the dependent variable (CAF measures). These correlations were calculated for the whole group of participants as well as for each group of participants, with separate analyses being conducted for the complex and the simple task. In addition, we also performed several regressions, which included three predictors: (a) working memory, (b) L2 proficiency, and (c) the interactions between working memory and L2 proficiency. The dependent variables included within the regressions were accuracy, lexical complexity, and syntactic complexity. We ran a separate multiple regression analysis for each of the dependent variables (for each dimension of the CAF measures) and performed separate regressions for the simple and complex tasks.

Results

Table 2 presents the descriptive statistics for the OPT (measure of proficiency as a continuous variable) and WM test for the participants ($n = 76$). Pearson product-moment correlations showed a small negative correlation between OPT and WM, which was nonsignificant ($r = -.133$).

Tables 3 and 4 display the descriptive statistics for the measures of L2 writing performance ($n = 76$) in the simple and complex tasks accordingly.

As can be seen from the mean values, the participants obtained rather similar values across the simple and complex task conditions on all measures, except for the measure of nominal complexity (ratio of complex nominals per clause), which was significantly higher in the simple task condition ($M = 1.05$) as compared with the complex task ($M = .54$), according to the paired samples t test, $t = -6.260$, $df = 75$, 95% CI $[-0.684, 0.354]$, $p = .001$.

Table 2. Descriptive statistics for OPT (proficiency) and WM test ($n = 76$).

Variables	Mean	SD	Min	Max	Skewness (SE)	Kurtosis (SE)
OPT	77.48	9.59	45.00	95.00	-.626 (.27)	.673 (.54)
WM test	1.03	.75	.30	3.26	.85 (.32)	.118 (.63)

Note. OPT = Oxford Placement Test; WM test = working memory test.

Table 3. Descriptive statistics for L2 writing performance (CAF) in the simple task ($n = 76$).

CAF variables	Mean (SD)	Min-Max	Skewness (SE)	Kurtosis (SE)
Ratio of errors	10.86 (4.83)	2.37–24.79	.822 (.276)	.497(.545)
Lexical density	.43 (.03)	.36–.52	.360 (.276)	.340 (.545)
Lexical variety (UBER)	13.73 (5.20)	5.40–23.19	-1.743(.276)	2.813 (.545)
Lexical sophistication	.15 (.03)	.03–.25	-.076 (.276)	-.101 (.545)
Length of T-unit	21.38 (5.63)	11.62–34.33	.373 (.276)	-.411 (.545)
Coordinate clause per clause	.22 (.14)	.03–.80	1.110 (.276)	2.633 (.545)
Dependent clause per clause	.44 (.14)	.07–.71	-.337 (.276)	-.384(.545)
Complex nominal per clause	1.05 (.38)	.44–2.50	1.818 (.276)	4.376 (.545)
Words per minute	12.45 (3.44)	4.53–20.31	.133 (.276)	-.478 (.545)
Total number of words	196 (72)	64–385	.622 (.276)	.079 (.545)

Table 4. Descriptive statistics for L2 writing performance (CAF) in the complex task ($n = 76$).

CAF variables	Mean (SD)	Min-Max	Skewness (SE)	Kurtosis (SE)
Ratio of errors	10.32 (5.50)	1.26–29.17	1.223 (.276)	1.469 (.545)
Lexical density	.44 (.02)	.37–.52	–.062 (.276)	.304 (.545)
Lexical variety (UBER)	14.90 (3.00)	.50–18.50	–3.259(.276)	14.82 (.545)
Lexical sophistication	.13 (.06)	.06–.27	–1.033 (.276)	.307 (.545)
Length of T-unit	22.30 (5.93)	10.49–41.00	.812 (.276)	.889 (.545)
Coordinate clause per clause	.24 (.14)	.14–.86	1.119 (.276)	3.088 (.545)
Dependent clause per clause	.43 (.13)	.13–6.7	–.470 (.276)	.582(.545)
Complex nominal per clause	.54 (.53)	.01–2.29	2.044 (.276)	6.898 (.545)
Words per minute	12.49 (.43)	6.26–24.44	.589 (.276)	.648 (.545)
Total number of words	223 (77)	74–625	2.051 (.276)	8.885(.545)

Table 5. Pearson correlations between the CAF measures of L2 writing production, OPT (proficiency) and WM test ($n = 76$).

CAF variables	Participants ($n = 76$)			
	Simple task		Complex task	
	OPT	WM test	OPT	WM test
Ratio of errors	–.573**	–.046	–.696**	.067
Lexical density	.166	–.154	.335*	.001
Lexical variety (UBER)	–.079	.071	.010	–.128
Lexical sophistication	.261*	–.214	–.065	.050
Length of T-unit	–.104	–.006	–.065	–.011
Coordinate clause per clause	–.078	–.061	.169	.007
Dependent clause per clause	–.032	.154	–.202	.096
Complex nominal per clause	–.080	.131	–.312**	.153
Words per minute	.251*	–.169	.245*	–.115
Total number of words	.295*	.011	.204	–.040

Note. Two-tailed tests.

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

In what follows we report the results according to the research questions guiding our study. Our first research question asked about the implication of WM in written language production. Table 5 summarizes Pearson product-moment correlations among the OPT scores, WM test, and the CAF variables.

The results of the correlations showed that the participants' level of L2 proficiency (as measured by the OPT), not WM, was the measure that correlated the most with L2 writing performance indices. Concerning the correlations between OPT and CAF measures, the most consistent results were obtained for the measures of accuracy (ratio of errors per 100 words) and fluency (words per minute). Specifically, strong negative correlations of $r = -.573$ and $r = -.696$ were observed between OPT and error rates in the simple and complex tasks. Additionally, there was a small positive correlation between OPT and the measure of fluency (words per minute; $r = .251$ and $r = .245$ in simple and complex task conditions, respectively).

There were no significant correlations between WM test and CAF measures of L2 writing production. However, it is worth noting that for error ratio, lexical density, lexical variety, lexical sophistication, coordination, and total number of words, the size and nature of the correlations with WM score were different in the simple and complex tasks (see Table 5).

Our second and third research questions asked about potential interactions between WM and task complexity (RQ2) and WM and L2 proficiency (RQ3). Thus, on the basis of the results obtained in the correlational analyses, we performed a series of multiple regressions in which the dependent variables were the CAF measures for which significant correlations had been obtained and the predictors were OPT scores (L2 proficiency), WM test score, and the interaction measure between WM and OPT scores. Separate regressions were performed for the CAF variables in the simple and complex task conditions. In an effort to control the interrelationships among variables, a test of multicollinearity was conducted. Resulting VIF values were all under 2, implying little threat of multicollinearity in the regression analyses. We first explored the scores for the OPT in the regression analysis, and then we entered the scores for WM and the interaction variable between OPT and WM to explore whether these additional variables contributed significantly to the predictive capacity of the model.

In the first step of the regression analysis with OPT as a predictor of the ratio of errors in the simple task, the model was significant, $F(1, 74) = 14.057, p \leq .001$, with the OPT explaining 21% of variance in the ratio of errors ($\beta = -.46, p \leq .001$; see Table 6). However, in a second step, when WM score was added, the model lost its significance ($p = .383$) and, accordingly, there were no noticeable changes in the explained variance ($\Delta R^2 = .012$); the significance of the model was at $p = .501$ when the interaction between OPT and WM was added as another predictor. A similar pattern of results was obtained for the ratio of errors in the complex task condition, in which the model was significant with the OPT as a single predictor, $F(1, 74) = 37.483, p \leq .001$, with the OPT explaining 41% of variance in the ratio of errors ($\beta = -.64, p \leq .001$; see Table 6). The addition of WM and the interaction variable between WM and OPT did not produce noticeable changes in the variance explained, with the model losing its significance.

Hierarchical multiple regression analysis was also performed to analyze the potential contribution of OPT, WM, and the interaction between OPT and WM to the measures of fluency (words per minute, total number of words written) and to the measures of lexical density, lexical sophistication, and nominal complexity in simple and complex tasks. In all these models, the OPT scores appeared as a single significant predictor, with the model losing its significance when WM and the interaction between OPT and WM scores were added as additional predictors. Thus, acting as a sole predictor, the OPT scores explained 6% of variance in the measure of words per minute (fluency) in the simple task, $F(1, 74) = 4.969, \beta = .25, p \leq .05$; similarly, 6% of variance in words per minute was explained by the OPT in the complex task, $F(1, 74) = 4.735, \beta = .24, p \leq .05$. The OPT scores also explained 8% of variance in the number of words (fluency) in the simple task, $F(1, 74) = 7.037, \beta = .29, p \leq .01$; but the model was not significant in the complex task ($p = .07$; see Table 7).

In the simple task, the OPT also correlated positively with lexical sophistication. The regression analysis showed that the OPT scores explained 7% of variance of lexical sophistication in the simple task, $F(1, 74) = 5.414, \beta = .26, p \leq .05$, whereas the model was not significant for lexical sophistication in the complex task ($p = .57$; see Table 8).

Table 6. Regression models explaining ratio of errors in simple and complex tasks with OPT as a predictor

Dependent variable	<i>R</i>	<i>R</i> ²	Adjusted <i>R</i> ²	SE of estimate	<i>p</i>
Ratio of errors: simple task	.48	.23	.21	3.89	.004
Ratio of errors: complex task	.65	.42	.41	4.19	≤.001

Table 7. Regression models explaining words per minute and the total number of words (measures of fluency) in the simple and complex tasks with OPT as a predictor

Dependent variable	<i>R</i>	<i>R</i> ²	Adjusted <i>R</i> ²	<i>SE</i> of estimate	<i>p</i>
Words per minute: simple task	.25	.08	.06	3.35	.029
Words per minute: complex task	.24	.07	.06	3.08	.033
Number of words: simple task	.29	.09	.08	69.59	.01
Number of words: complex task	.20	.04	.03	76.23	.07

Table 8. Regression models explaining lexical sophistication in simple and complex tasks with OPT as a predictor

Dependent variable	<i>R</i>	<i>R</i> ²	Adjusted <i>R</i> ²	<i>SE</i> of estimate	<i>p</i>
Lex. Sophistication: simple task	.26	.08	.07	.03	.023
Lex. Sophistication: complex task	.06	.01	.01	.06	.576

Table 9. Regression models explaining lexical density in simple and complex tasks with OPT as a predictor.

Dependent variable	<i>R</i>	<i>R</i> ²	Adjusted <i>R</i> ²	<i>SE</i> of estimate	<i>p</i>
Lex. density: simple task	.16	.02	.01	.03	.152
Lex. density: complex task	.33	.09	.11	.02	.003

Table 10. Regression models explaining nominal complexity in simple and complex tasks with OPT as a predictor.

Dependent variable	<i>R</i>	<i>R</i> ²	Adjusted <i>R</i> ²	<i>SE</i> of estimate	<i>p</i>
Nominal complexity: simple task	.08	.01	.00	.38	.489
Nominal complexity: complex task	.31	.09	.08	.50	.006

For some measures, the role of the OPT scores was more prominent in the complex task condition: the OPT scores explained 11% of variance of lexical density in the complex task, $F(1, 74) = 9.344$, $\beta = .33$, $p \leq .01$, whereas the model was not significant in the simple task ($p = .152$; see Table 9).

Also, the regression analysis showed a negative relationship between the OPT scores and nominal complexity in the complex task, $F(1, 74) = 7.968$, $\beta = -.31$, $p \leq .01$, whereas the model was not significant in the simple task ($p = .48$; see Table 10).

Discussion

Our study investigated how WM may be implicated in L2 written performance and whether L2 proficiency and task complexity moderate any WM effects. As a global summary, our findings point to some independent and interactive effects of the predictor variables. Thus, L2 proficiency emerged as the sole significant predictor of L2 writing performance at both levels of task complexity. In contrast, and contrary to our expectations, no significant WM effects on text characteristics (in terms of CAF measures) were observed. In terms of interactive effects, no significant interaction between WM and task complexity was found, whereas our data can be interpreted as suggesting that the role of proficiency in L2 writing may vary depending on task

complexity. In this sense, L2 proficiency explained 21% of variance in accuracy in the simple task but 41% of variance in the complex task. Additionally, depending on the level of task complexity, proficiency played a different role for some measures of complexity and the total number of words produced. Thus, higher proficiency correlated positively with higher number of words and higher lexical density and lexical sophistication only in the complex task, whereas the opposite tendency was observed for nominal complexity.

In what follows we interpret these findings, separating the independent and interactive effects (or lack of) observed.

Working memory effects on L2 written production: Independent effects of proficiency and task complexity

Our first RQ asked the extent to which working memory affects L2 written performance, the latter operationalized in our study in terms of complexity, accuracy, and fluency indices.

The correlational data showed no significant effects of WM on the characteristics of the texts produced. The observed lack of WM effects is not consistent with most previous research reporting a connection between WM and written performance (Adams & Guillot, 2008; Baoshu & Chuanbi, 2015; Bergsleithner, 2010; Kormos & Sáfár, 2008; Mavrou, 2020; Michel *et al.*, 2019; Mujtaba *et al.*, 2021; Peng *et al.*, 2022; Révész *et al.*, 2017; Vasylets & Marín, 2021; Zabihi, 2018; Zalbidea, 2017), but it is a finding in line with the small number of previous studies that found insignificant or null WM effects (Cho, 2018; Kim *et al.*, 2021; Lu, 2015). One possible explanation could be related to the WM instrument used in the research design. Our research measured WM via the online *n*-back test, an instrument used primarily in neuroscience and psychology research, not in L2 writing research. Yet, it is similar to the running span task used in Kim *et al.* (2021), a study that also reported no WM effects on production. Of relevance, the existing L2 writing literature reporting WM effects has used a range of diverse WM measures, including nonword span tests (Kormos & Sáfár, 2008), visual forward and backward digit span tests (Kormos & Sáfár, 2008; Michel *et al.*, 2019), and complex span tests (Vasylets & Marín, 2021; Zalbidea, 2017). The rationale for positing this possible methodological explanation related to how WM is tested gains additional weight when we consider that some studies that reported no WM effects on production (Cho, 2018; Lu, 2015) used the same WM test—namely, an operation span test. Interestingly, this WM test was only used in two of the studies reporting WM effects on production (Mujtaba *et al.*, 2021; Zalbidea, 2017). In other words, the WM tests used in studies that have reported no WM effects (including ours) were different from those tests used in the majority of studies reporting WM effects on all or some CAF indices. Accordingly, it might be speculated that the WM test used in L2 writing research might constitute a crucial methodological issue that ought to be seriously considered in future L2 writing research, echoing a general call in the L2-WM literature to make methodological issues more central in WM studies in the domain of writing and WM research globally (see Shin & Hu, 2022). In the case of L2 writing, Kim *et al.* (2021) raise the issue of the relevance of using language-independent, nonverbal WM tests and of the use of simple and complex span tests. From a more global perspective, Shin and Hu, make the following recommendation, which would be worth applying to research on WM effects on writing: “perhaps we can start small by making it a norm to specifically and explicitly state the operational construct of WM and components (e.g., storage, processing,

executive WM) targeted by the WM task used to help research consumers and researchers differentiate between and compare across studies” (p. 738).

Perhaps future work trying to ascertain WM effects on written performance could engage in partial replication of extant studies, conducting research with similar groups of participants performing the same tasks and under the same task implementation conditions as in the original study but varying the WM test used in the study being replicated. Additionally, WM L2 writing studies could use more than one WM task (as done, for instance, in Mavrou, 2020), which has been suggested for global WM research in order to “obtain an average of scores to estimate WM capacity” (Shin & Hu, 2022, p. 739). This way a more nuanced understanding of why and how WM effects (or lack thereof) vary as a function of how WM capacity is tested could be gained.

An additional potential explanation of our findings regarding the lack of WM effects could be related to the participants themselves. Coincidentally, similar to our own study, the three previous studies reporting no WM effects were conducted with university students: The participants in Cho’s (2018) study were 39 Korean EFL university students majoring in English language and literature (as was also the case with the participants in our own study). Kim et al. (2021) conducted their study with 100 undergraduate students from various countries, and Lu’s (2015) participants were 136 undergraduate and graduate students from various disciplines. Yet, academic background does not appear as a fully convincing explanation given that some previous work reporting WM effects had also been conducted with university students (e.g., Vasylets & Marin, 2021; Zalbidea, 2017), in some cases in a foreign language degree program (e.g., Zalbidea, 2017), as was the case with our own participants. As suggested by Sanz (personal communication), a plausible explanation might be that college-age students, unlike children or the elderly L2 user, are in their cognitive prime, and, unless the task is very taxing, WM constraints do not come to the surface. Additionally, literacy-related skills might also be involved, as Kim et al. (2021) found in their own study. These observations and speculations point to the relevance of, first, expanding populations in future research on WM effects on L2 writing and, second, to the relevance of inspecting more closely potential interactions of cognitive maturity and literacy-related skills and resources when making decisions on the tasks used (and how taxing they may be) in WM studies with adults.

In contrast to the virtually null effects of WM on L2 performance, we found that L2 proficiency had a significant positive effect on the quality of L2 performance, especially in the areas of accuracy and fluency. The pattern of findings was particularly revealing for accuracy, as we consistently found significant negative correlations between the ratio of errors per 100 word and L2 proficiency across the two task-complexity conditions. The results from the correlations were further confirmed by the regression analysis, which showed that L2 proficiency was a sole significant predictor of L2 writing accuracy. Thus, L2 proficiency accounted for 21% and 41% of variance in the simple and complex task conditions, respectively. These findings, which are in line with previous empirical research reporting positive relationship between L2 proficiency and writing accuracy (Kim et al., 2016; Wolfe-Quintero et al., 1998), can be explained by considering that accuracy in language use can result from the interaction of various sources including the degree of accuracy of the linguistic representations in learners’ interlanguage or the strength of the competing representations (Wolfe-Quintero et al., 1998). It would be plausible to consider that higher L2 proficiency can contribute to each source of writing accuracy. Thus, with higher levels of L2 knowledge, the repertoire of internal linguistic representations is broader, and a higher correspondence of the internalized linguistic items with the standardized linguistic items in the L2

exists. As L2 proficiency increases, L2 users are also better equipped to suppress erroneous linguistic representations that can constitute a source of errors in L2 production.

Our results also revealed a positive contribution of L2 proficiency to L2 writing fluency, with L2 proficiency explaining 6% of variance in fluency in both simple and complex task conditions. These findings are in line with numerous previous studies reporting a positive connection between fluency of language production and overall L2 proficiency (Baker-Smemoe *et al.*, 2014; de Jong *et al.*, 2015; Larsen-Freeman, 2009; Segalowitz, 2010). Writing fluency is a multidimensional construct and involves the ability to produce written language rapidly, appropriately, creatively, and coherently (Abdel Latif, 2013; Wolfe-Quintero *et al.*, 1998). In our study, we employed the measure of words per minute, which taps into the speed dimension of writing fluency. Positive links between L2 proficiency and speed fluency of writing production can be explained by the fact that at higher levels of proficiency, learners' L2 knowledge is characterized by a higher level of proceduralization (Schmidt, 1992), resulting in more efficient and rapid retrieval of linguistic representations during language production.

Regarding task complexity effects, results showed that the participants performed almost identically across the simple and complex task conditions in most measures of L2 written production. The only significant difference was the measure of nominal complexity, which appeared significantly higher in the simple task condition, in line with previous research that found syntactic complexity to be the one measure that was not affected by an increase in task complexity (Kuiken *et al.*, 2005; Kuiken & Vedder, 2007; Michel *et al.*, 2007).

Interactive effects of working memory, proficiency, and task complexity

Our second and third research questions asked about potential interactions between WM and proficiency, on one hand, and WM and task complexity, on the other. To this end, we explored correlations between WM and proficiency, WM and task complexity, and L2 proficiency and task complexity.

As noted above, we found significant effects of L2 knowledge on L2 writing performance, as compared with the lack of significant WM effects. In relation to our second research question, our results also showed an absence of interactive effects between L2 proficiency and WM in relation to the quality of L2 writing. These null interactive effects are in line with Lu (2015) and also partially confirm findings in Vasylets and Marín (2021), who found interactive effects of WM and L2 proficiency only for the selected dimensions of performance (in particular, accuracy and lexical sophistication), whereas there were null effects for the dimensions of syntactic complexity, lexical diversity, and fluency. These findings confirm the complex pattern of the involvement of cognitive resources in L2 production (Williams, 2015) as well as the moderating role of additional variables' effects mentioned in the background to the study. In this respect, the complexity of the influences of WM can be due to multiple factors, as well as their interactions, which can determine the pattern of involvement of cognitive resources during the completion of an L2 task. This resonates with the theories of cognitive psychology that posit a variety of scenarios for the interactive effects between WM and knowledge. Thus, the theory of compensation views WM as a compensatory mechanism at low levels of knowledge (Ackerman, 1988) and consequently predicts greater involvement of WM at lower levels of knowledge. An alternative view is advanced in the rich-get-richer hypothesis (Hambrick & Engle, 2002),

which posits greater prominence of WM at higher levels of knowledge. In this view, WM is purported to function as a facilitating mechanism, a conduit of knowledge, rather than a mechanism of compensation. In the realm of SLA, empirical findings on oral production provide evidence for both of these differing predictions. For example, findings in Serafini and Sanz (2016) showed that WM played a positive role in the morphosyntactic development of L2 learners only at low levels of L2 proficiency as measured by grammaticality judgement tests (GJTs), whereas Gilabert and Muñoz (2010) found that the involvement of WM in L2 oral production was evident only at higher levels of proficiency. In addition to these two opposing scenarios, which contemplate WM effects at low or high levels of proficiency, we can also propose a middle-ground scenario of a null interaction of WM and L2 knowledge/proficiency, as found in some previous L2 writing studies (e.g., Lu, 2015) and in our own. Such null interaction could be due to task-related factors. Thus, we could suggest that the higher control of time and availability of planning time inherent in written production (see below for further elaboration of time-on-task considerations) created propitious conditions in which the task used in our study, which provided clear instructions and a visual prompt, allowed our participants to rely solely on their linguistic knowledge (and probably their literacy resources, although we did not measure this), without a detectable involvement of WM resources. As suggested by Sanz (personal communication), using a WM test that measures executive control would have produced different results—once again a reminder of the relevance of putting methodological considerations at central stage in future WM-L2 writing studies.

The speculation about potential effects of task-related considerations is further reinforced when we consider the data on the interaction between WM and task complexity. Initially, in accordance with Robinson's (2005, 2011) predictions regarding the likelihood of a more prominent role of IDs within complex tasks and taking into consideration the extant literature on the interaction between WM and task complexity (e.g., Kormos & Trebits, 2011; Michel et al., 2019; Zalbidea, 2017), we anticipated that WM would play a more significant role in the complex task due to its higher cognitive demands and its increased problem-solving nature. Nevertheless, in addition to the fact that WM did not emerge as a significant predictor of L2 written production, our findings also revealed no significant interaction between WM and task complexity. As advanced above in the case of WM–proficiency interactions, we would speculate that the lack of interaction between WM and task complexity could be a function of the task itself and task implementation conditions in our study. As for the task itself, it may not have been demanding enough. In this respect, McCormick and Sanz (2022) argue that the role of WM “reveals itself empirically only when the task pushes the learners to their cognitive limits” (p. 586). When discussing the role of proficiency, they argue for the relevance of increasing the challenge tasks pose “in order to see a differential role at higher levels of proficiency” (p. 583) and thus conclude that “studies with both multiple proficiency levels and multiple tasks of increasing complexity may prove useful to further probe WMC advantages and the dynamic nature of WM” (p. 586). This is precisely what we tried to do in our study, but given that our participants were allowed ample time to complete their complex and simple versions of the task (50 min), potential WM effects on the allocation of attentional resources, orchestration of writing processes, and shifting between processes in the more- and less-complex task could have been neutralized by the extended task-time conditions in which the participants completed their writing. In this respect, time on task might be a crucial consideration when ascertaining the relevance of WM effects on language use more generally, which would call for some caution in generalizing existing findings in the SLA WM literature

from learning through input processing to learning by output production and, within the latter, WM effects across language modalities. This is so because, as repeatedly discussed in theoretical accounts of writing as a site for language learning (e.g., Manchón, 2023; Manchón & Williams, 2016; Williams, 2012), the time pace that characterizes writing (which, with the exception of some forms of online written interactions, takes place offline, in contrast to the on-line nature of oral communication) allows for L2 writers to be more in control of their attentional resources (with potentially less involvement of WM) and likely to focus on linguistic concerns during completing writing tasks.

It is also relevant to note that time on task alone does not explain previous conflicting findings on WM effects or lack thereof in the case of writing (regardless of whether or not the study considered the moderating effect of task complexity). Just as examples, task time varied in the three studies that reported no WM effects on written performance: 30 min in the case of Lu (2015), 25 min in Kim *et al.* (2021), and 10 to 20 min in Cho (2018). Nevertheless, the same amount of time on task was reported in studies showing positive WM effects: 15 to 30 min in Vasylets and Marín's (2021) study and 20 min in Mavrou's (2020) work. It is probably the consideration of the combination of time on task and the nature of the task itself that is relevant for, echoing McCormick and Sanz (2022), WM effects to reveal themselves empirically. Applied to our own study, the 50 min established in our task instructions appear to have provided our participants with ample time (which they did not have to use in total) to attend the demands of the tasks at hand (and do so in on the basis of their L2 knowledge resources and probably their literacy skills, as discussed above) regardless of the inherent complexity of the task. In support of this interpretation, CAF indices hardly varied across the complex and simple versions of the task, with the only exception of "complex nominal per clause" and total number of words, which coincidentally were the two indices with the highest *SD* (see also Tables 3 and 4 in the Results section).

The larger number of words in the complex task could easily be explained by task-related conditions, especially the greater number of elements the participants had to account for in the complex condition. Recall that the task instructions asked participants to explain (a) what actions they would take in order to save as many people as possible from the burning building, (b) in what order they would rescue these people, and (c) why they would take these actions. The complex task entailed an increase in reasoning demands because of the elements included (more people to be rescued, more fire within the building, less emergency services available, etc.), which could easily lead to requiring extra words to complete the task. Writing longer texts was also facilitated by the ample time-on-task conditions in our study.

Conclusion

Our study set out to examine the independent effects of working memory and the interactive effects of working memory, L2 proficiency, and task complexity on L2 written performance. The results of the correlations distinctively showed that L2 proficiency emerged as a stronger predictor of L2 writing performance than the L2 writers' cognitive ability (WM). Thus, the results indicate that, for the population under study, when task complexity is operationalized in terms of reasoning demands and when WM is measured by the *n*-back test, the amount of L2 knowledge the L2 writer has exerted a stronger influence on their L2 written production than their WM capacity.

We believe that our study contributes to previous work on WM effects in writing in two ways. First, the research reported in this paper adds to previous work on cognitive IDs in conjunction with L2 writing by not only focusing on the effects of WM on L2 written production but also by combining in one and the same study an inquiry into the potential interactions between learner-related variables (working memory and L2 proficiency) and task-related variables (task complexity) that have hitherto been addressed separately. This more complex design allowed us to shed a stronger light on the independent and interactive effects on WM, L2 proficiency, and task complexity. Second, we would also argue that the study may constitute a contribution to important methodological considerations in future work, especially regarding time-on-task conditions and the way in which WM is tested. As mentioned in the Discussion, future studies on the implication of WM in written output should test potential effects of task implementation conditions, take principled decisions as to which WM test to use, and even partially replicate previous work varying task time and the way in which WM is tested.

In terms of wider implication for SLA studies on WM effects, our results call for caution in extrapolating available findings on WM effects on learning and language use to the writing domain on account of, at a minimum, the inherent problem-solving and time nature of writing. More precisely, it is relevant to be mindful that, as mentioned in previous sections, writing entails knowledge and skills to orchestrate the demands of and writing processes involved in text production. Yet, the pace of writing (and resulting expanded time-on-task conditions) may moderate (and even neutralize in the case of certain tasks and for specific groups of L2 users) the implication of WM in the activation and coordination of writing processes and resulting allocation of attentional resources to all the dimension of composing, crucially including writers' own decisions about language-related concerns to be addressed and capability to successfully address such concerns.

Despite this potential contribution of our research, the study is not without its limitations, particularly regarding participants. Thus, given the number of variables included, the sample size was relatively small. In addition, the participants were all university-level students from a language and linguistics undergraduate degree, thus limiting results to a very specific profile of L2 users in terms of age range, academic background, and literacy skills and resources. The participants' level of L2 proficiency may have also facilitated their completion of the research tasks on the basis of their own linguistic resources, thus limiting potential WM effects on their L2 written texts. It is also essential to keep in mind the instrument used to measure WM. As previously mentioned, measures differ not only in different research fields but also within SLA research itself and certainly in studies exploring WM and L2 writing. Further research would benefit from exploring these instruments in depth to shed light on their validity to explore WM effects on L2 writing. We must also admit limitations in the way fluency was assessed. In this study, we employed the total number of words and words per minute as measures of fluency. Although Wolfe-Quintero et al. (1998) argued that the number of words was the fluency measure that distinguished best between writers at different proficiency levels, they also admitted that this measure might not be entirely reliable because of the mixed results obtained in some studies. Moreover, we did not employ any process-based measure of fluency (e.g., length of text between pauses) in our study. Future research should opt for multidimensional assessment of fluency, combining speed measures (words/syllables per minute), product-based measures (number of words), and also process-based measures (see Kim et al., 2021; Révész et al., 2017 and Torres in this special issue).

Despite these limitations, we would like to suggest that our study contributes one additional piece (perhaps more relevant from a research methodological perspective as concerns such crucial dimensions as the complexity—or lack thereof—research designs, time-on-task considerations, and actual measurement of WM capacity) to the growing research interested in testing theoretical predictions on the implication of WM in written language use in an additional language.

Acknowledgments. The global program of research on L2 writing within which this article is situated was financed by the Spanish Research Agency (Research Grant PID2019-106091GB-I00 funded by MCIN/AEI/10.13039/501100011033 and Pre-Doctoral Grant BES-2017-081873) as well as The Seneca Foundation (Research Grant 20832/PI/18).

Competing interest. The author(s) declare none

References

- Abdel Latif, M. M. (2013). What do we mean by writing fluency and how can it be validly measured? *Applied Linguistics*, 34, 99–105. <https://doi.org/10.1093/applin/ams073>.
- Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology*, 117, 288–318. <https://doi.org/10.1037/0096-3445.117.3.288>
- Adams, A. M., & Guillot, K. (2008). Working memory and writing in bilingual students. *International Journal of Applied Linguistics*, 156, 13–28. <https://doi.org/10.2143/ITL.156.0.2034417>
- Allen, D. (1992). *Oxford Placement Test 1: Marking kit with user's guide and diagnostic key*. Oxford University Press.
- Awwad, A., & Tavakoli, P. (2022). Task complexity, language proficiency and working memory: Interaction effects on second language speech performance. *International Review of Applied Linguistics in Language Teaching*, 60, 169–196. <https://doi.org/10.1515/iral-2018-0378>
- Baddeley, A. (1986). *Working memory*. Clarendon Press/Oxford University Press. <https://doi.org/10.1002/acp.2350020209>
- Baddeley, A. (2015). Working memory in second language learning. In Z. E. Wen, M. B. Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 17–28). Multilingual Matters.
- Baker-Smoeve, W., Dewey, D. P., Bown, J., & Martinsen, R. A. (2014). Does measuring L2 utterance fluency equal measuring overall L2 proficiency? Evidence from five languages. *Foreign Language Annals*, 47, 707–728.
- Baoshu, Y., & Chuanbi, N. (2015). Planning and working memory effects on L2 performance in Chinese EFL learners' argumentative writing. *Indonesian Journal of Applied Linguistics*, 5, 43–51. <https://doi.org/10.17509/ijal.v5i1.830>
- Baoshu, Y., & Luo, S. (2012). The effect of working memory capacity on written language production of second language learners. *Foreign Language Teaching and Research*, 44, 536–546.
- Barkaoui, K. (2016). What and when second-language learners revise when responding to timed writing tasks on the computer: The roles of task type, second language proficiency, and keyboarding skills. *The Modern Language Review*, 100, 320–340. <https://doi.org/10.1111/modl.12316>
- Barkaoui, K. (2019). What can L2 writers' pausing behavior tell us about their L2 writing processes? *Studies in Second Language Acquisition*, 41, 529–554. <https://doi.org/10.1017/S027226311900010X>
- Bergsleithner, J. M. (2010). Working memory capacity and L2 writing performance. *Ciências & Cognição*, 15, 2–20.
- Cho, M. (2018). Task complexity, modality, and working memory in L2 task performance. *System*, 72, 85–98. <https://doi.org/10.1016/j.system.2017.10.010>
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12, 769–786. <https://doi.org/10.3758/BF03196772>

- Coulacoglou, C., & Saklofske, D. H. (2017). Executive function, theory of mind, and adaptive behavior. In C. Coulacoglou & D. H. Saklofske (Eds.), *Psychometrics and psychological assessment* (pp. 91–130). Academic Press. <https://doi.org/10.1016/B978-0-12-802219-1.00005-5>
- de Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, 36, 223–243. <https://doi.org/10.1017/S0142716413000210>
- Dugast, D. (1979). *Vocabulaire et stylistique. I The 'atre et ' dialogue* [Vocabulary and style. Vol. 1 Theatre and dialogue]. Slatkine-Champion.
- Gánem-Gutiérrez, G. A., & Gilmore, A. (2018). Tracking the real-time evolution of a writing event: Second language writers at different proficiency levels. *Language Learning*, 68, 469–506. <https://doi.org/10.1111/lang.12280>
- Gajewski, P. D., Hanisch, E., Falkenstein, M., Thönes, S., & Wascher, E. (2018). What does the *n*-back task measure as we get older? Relations between working-memory measures and other cognitive functions across the lifespan. *Frontiers in Psychology*, 9, 1–17. <https://doi.org/10.3389/fpsyg.2018.02208>
- Gilbert, R. (2005). Evaluating the use of multiple sources and methods in needs analysis: A case study of journalists in the Autonomous Community of Catalonia (Spain). In M. H. Long (Ed.), *Second language needs analysis* (pp. 182–199). Cambridge University Press.
- Gilbert, R. (2007). Effects of manipulating task complexity on self-repairs during L2 oral production. *IRAL—International Review of Applied Linguistics in Language Teaching*, 45, 215–240. <https://doi.org/10.1515/iral.2007.010>
- Gilbert, R., & Muñoz, C. (2010). Differences in attainment and performance in a foreign language: The role of working memory capacity. *International Journal of English Studies*, 10(1), 19–42. <https://doi.org/10.6018/ijes/2010/1/113961>
- Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. *Cognitive Psychology*, 44, 339–387. <https://doi.org/10.1006/cogp.2001.0769>
- Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 1–27). Lawrence Erlbaum.
- Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication*, 29, 369–388. <https://doi.org/10.1177/0741088312451260>
- Ishikawa, T. (2007). The effect of increasing task complexity along the [±Here-and-Now] dimension on L2 written narrative discourse. In M. P. Garcia Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 136–156). Multilingual Matters.
- Jaeggi, S. M., Buschkuhl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the *n*-back task as a working memory measure. *Memory*, 18, 394–412. <https://doi.org/10.1080/09658211003702171>
- Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 21–46). Oxford University Press.
- Kellogg, R. T. (1996). A model of working memory in writing. In M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 57–72). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9780203811122>
- Kellogg, R. T. (2001). Competition for working memory among writing processes. *American Journal of Psychology*, 114, 175–192. <https://doi.org/10.2307/1423513>
- Kim, Y., Nam, J., & Lee, S. Y. (2016). Correlation of proficiency with complexity, accuracy, and fluency in spoken and written production: Evidence from L2 Korean. *Journal of the National Council of Less Commonly Taught Languages*, 19, 147–181.
- Kim, M., Tian, Y., & Crossley, S. (2021). Exploring the relationship among cognitive and linguistic resources, writing processes, and written products in second language writing. *Journal of Second Language Writing*, 58, Article 100824. <https://doi.org/10.1016/j.jslw.2021.100824>
- Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing*, 21, 390–403. <https://doi.org/10.1016/j.jslw.2012.09.003>
- Kormos, J., & Sáfár, A. (2008). Phonological short-term memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, 11, 261–271. <https://doi.org/10.1017/S1366728908003416>

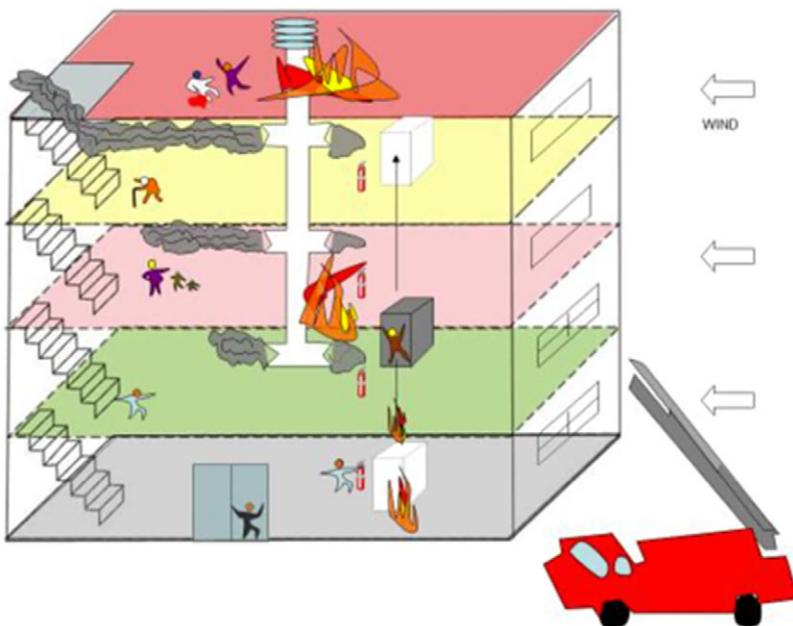
- Kormos, J., & Trebits, A. (2011). Working memory capacity and narrative task performance. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 267–286). John Benjamins. <https://doi.org/10.1075/tblt.2.17ch10>
- Kuiken, F. & Vedder, I. (2007). Task complexity and measures of linguistic performance in L2 writing. *International Review of Applied Linguistics in Language Teaching (IRAL)*, 45, 261–284. <https://doi.org/10.1515/iral.2007.012>
- 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., 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>
- Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30, 579–589. <https://doi.org/10.1093/applin/amp043>
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus*. Longman.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15, 474–496. <https://doi.org/10.1075/ijcl.15.4.02lu>
- Lu, X. (2011). A corpus-based evaluation of grammatical complexity measures as indices of college level ESL writers' language development. *TESOL Quarterly*, 45, 36–62. <https://doi.org/10.5054/tq.2011.240859>
- Lu, Y. (2015). Working memory, cognitive resources and L2 writing performance. In A. Wen, Z. Mota, & M. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 175–188). Multilingual Matters. <https://doi.org/10.21832/9781783093595-014>
- Manchón, R. M. (2023). The psycholinguistics of L2 writing. In A. Godfroid & H. Hopp (Eds.), *The Routledge handbook of second language acquisition and psycholinguistics* (pp. 400–412). Routledge.
- Manchón, R. M., & Roca de Larios, J. (2007). On the temporal nature of planning in L1 and L2 composing: A study of foreign language writers. *Language Learning*, 57(4), 549–593. <https://doi.org/10.1111/j.1467-9922.2007.00428.x>
- Manchón, R. M., Roca de Larios, J., & Murphy, L. (2009). The temporal dimension and problem-solving nature of foreign language composing. Implications for theory. In R. M. Manchón (Ed.), *Writing in foreign language contexts: Learning, teaching and research* (pp. 102–124). Multilingual Matters. <https://doi.org/10.21832/9781847691859-008>
- Manchón, R. M., & Williams, J. (2016). L2 writing and SLA studies. In R. M. Manchón & P. K. Matsuda (Eds.), *The handbook of second and foreign language writing* (pp. 567–586). de Gruyter Mouton.
- Mavrou, I. (2020). Working memory, executive functions, and emotional intelligence in second language writing. *Journal of Second Language Writing*, 50, Article 100758. <https://doi.org/10.1016/j.jslw.2020.100758>
- McCormick, T., & Sanz, C. (2022). Working memory and L2 grammar learning among adults. In J. Schwieter & Z. Wen (Eds.), *The Cambridge handbook of working memory and language* (pp. 573–592). Cambridge University Press. doi:10.1017/9781108955638.032
- McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review*, 8, 299–325.
- Michel, M., Kormos, J., Brunfaut, T., & Ratajczak, M. (2019). The role of working memory in young second language learners' written performances. *Journal of Second Language Writing*, 45, 31–45. <https://doi.org/10.1016/j.jslw.2019.03.002>
- Michel, M. C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *International Review of Applied Linguistics in Language Teaching (IRAL)*, 45, 241–259. <https://doi.org/10.1515/iral.2007.011>
- Michel, M., Révész, A., Lu, X., Kourтали, N., Lee, M., & Borges, L. (2020). Investigating L2 writing processes across independent and integrated tasks: A mixed methods study. *Second Language Research*, 36, 307–334. <https://doi.org/10.1177/0267658320915501>
- Mujtaba, S. M., Kamyabi Gol, A., & Parkash, R. (2021). A study on the relationship between language aptitude, vocabulary size, working memory, and L2 writing accuracy. *Foreign Language Annals*, 54, 1059–1081. <https://doi.org/10.1111/flan.12584>

- Olive, T. (2011). Working memory in writing. In V. W. Berninger (Ed.), *Past, present, and future contributions of cognitive writing research to cognitive psychology* (pp. 485–506). Psychology Press. <https://doi.org/10.4324/9780203805312>
- Olive, T. (2022). How do novice and skilled writers engage working memory? In J. Schwieter & Z. Wen (Eds.), *The Cambridge handbook of working memory and language* (pp. 504–526). Cambridge University Press. doi:10.1017/9781108955638.028
- Peng, A., Orosco, M. J., Wang, H., Swanson, H. L., & Reed, D. K. (2022). Cognition and writing development in early adolescent English learners. *Journal of Educational Psychology*, 114, 1136–1155. <https://doi.org/10.1037/edu0000695>
- Read, J. (2000). *Assessing vocabulary*. Oxford University Press.
- Révész, A., Michel, M., & Gilabert, R. (2016). Measuring cognitive task demands using dual-task methodology, subjective self-ratings, and expert judgments. *Studies in Second Language Acquisition*, 38, 703–737. <https://doi.org/10.1017/S0272263115000339>
- Révész, A., Michel, M., & Lee, M. (2017). *Investigating IELTS Academic Writing Task 2: Relationships between cognitive writing processes, text quality, and working memory*. IELTS Research Report. The British Council.
- Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a compartmental framework. *Applied Linguistics*, 22, 27–57. <https://doi.org/10.1093/applin/22.1.27>
- Robinson, P. (2005). Cognitive complexity and task sequencing: A review of studies in a compartmental framework for second language task design. *International Review of Applied Linguistics*, 43, 1–32. <https://doi.org/10.1515/iral.2005.43.1.1>
- Robinson, P. (2011). 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* (3–38). John Benjamins. <https://doi.org/10.1075/tblt.2.05ch1>
- Roca de Larios, J., Manchón, R., Murphy, L., & Marin, J. (2008). The foreign language writer's strategic behavior in the allocation of time to writing processes. *Journal of Second Language Writing*, 17, 30–47. <https://doi.org/10.1016/j.jslw.2007.08.005>
- 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>
- Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, 14, 357–385. <https://doi.org/10.1017/S0272263100011189>
- Schoonen, R., Van Gelderen, A., Stoel, R. D., Hulstijn, J., & De Glopper, K. (2011). Modeling the development of L1 and EFL writing proficiency of secondary school students. *Language Learning*, 61, 31–79. <https://doi.org/10.1111/j.1467-9922.2010.00590.x>
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge. <https://doi.org/10.4324/9780203851357>
- Serafini, E. J., & Sanz, C. (2016). Evidence of the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition*, 38, 607–646. <https://doi.org/10.1017/S0272263115000327>
- Shin, J., & Hu, Y. (2022). A methodological synthesis of working memory tasks in L2 research. In J. Schwieter & Z. Wen (Eds.), *The Cambridge handbook of working memory and language* (pp. 722–746). Cambridge University Press. doi:10.1017/9781108955638.038
- Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, 42, 1096–1104. <https://doi.org/10.3758/BRM.42.4.1096>
- Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44, 24–31. <https://doi.org/10.1177/0098628316677643>
- Tillema, M. (2012). *Writing in first and second language. Empirical studies on text quality and writing processes* (Publication No. 229) [Doctoral dissertation, Netherlands National Graduate School of Linguistics] LOT Publications. <https://www.lotpublications.nl/writing-in-first-and-second-language-writing-in-first-and-second-language-empirical-studies-on-text-quality-and-writing-processes>
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37, 498–505. <https://doi.org/10.3758/BF03192720>
- University of Cambridge Local Examinations Syndicate (UCLES). (2001). *Oxford quick placement test*. Oxford University Press.
- Van Weijen, D. (2009). *Writing processes, text quality, and task effects. Empirical studies in first and second language writing* [Doctoral dissertation, Netherlands National Graduate School of Linguistics]

- (Publication No. 201). LOT Dissertation Series. <https://www.lotpublications.nl/writing-processes-text-quality-and-task-effects-writing-processes-text-quality-and-task-effects-empirical-studies-in-first-and-second-language-writing>
- Vasylets, O., Gilbert, R., & Manchón, R. M. (2017). The effects of mode and task complexity on second language production. *Language Learning*, 67, 394–430. <https://doi.org/10.1111/lang.12228>
- Vasylets, O., & Marin, J. (2021). The effects of working memory and L2 proficiency on L2 writing. *Journal of Second Language Writing*, 52, Article 100786. <https://doi.org/10.1016/j.jslw.2020.100786>
- Weigle, S. C. (2005). Second language writing expertise. In K. Johnson (Ed.), *Expertise in second language learning and teaching* (pp. 128–149). Palgrave Macmillan. https://doi.org/10.1057/9780230523470_7
- Williams, J. (2012). The potential role(s) of writing in second language development. *Journal of Second Language Writing*, 21, 321–331.
- Williams, J. (2015). Working memory in SLA research: Challenges and prospects. In Z. E. Wen, M. B. Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 301–308). Multilingual Matters. <https://doi.org/10.21832/9781783095735>
- Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity*. University of Hawai'i, Second Language Teaching & Curriculum Center.
- Zabihi, R. (2018). The role of cognitive and affective factors in measures of L2 writing. *Written Communication*, 35, 32–57. <https://doi.org/10.1177/0741088317735836>
- Zalbidea, J. (2017). “One task fits all?” The roles of task complexity, modality, and working memory capacity in L2 performance. *Modern Language Journal*, 101, 335–352. <https://doi.org/10.1111/modl.12389>

Appendix 1

Complex version of “Fire-Chief” task (Gilbert, 2005)

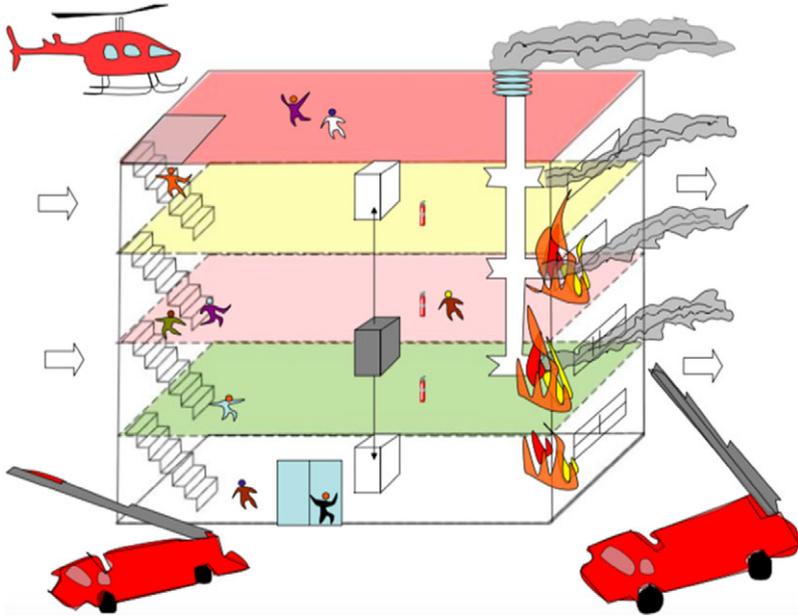


WRITING TASK

Instructions: Observe the fire chief image and **write a description**, considering all the different elements involved in the situation. More precisely, you have to explain (a) which action you would take to save as many people as possible, and (b) the sequence (i.e., order) in which you would take those actions. In both cases, you have to justify your choice of actions and their sequence. [In short, say what you would do, in which order, and why.]

Appendix 2

Simple version of “Fire-Chief” task (Gilabert, 2005)



WRITING TASK

Instructions: Observe the fire chief image and **write a description**, considering all the different elements involved in the situation. More precisely, you have to explain (a) which action you would take to save as many people as possible, and (b) the sequence (i.e., order) in which you would take those actions. In both cases, you have to justify your choice of actions and their sequence. [In short, say what you would do, in which order, and why.]

Cite this article: Manchón, R. M., McBride, S., Mellado Martínez, M. D. and Vasylets, O. (2023). Working memory, L2 proficiency, and task complexity: Independent and interactive effects on L2 written performance. *Studies in Second Language Acquisition*, 45: 737–764. <https://doi.org/10.1017/S0272263123000141>