Journal of Second Language Writing 52 (2021) 100786

Contents lists available at ScienceDirect

Journal of Second Language Writing

journal homepage: www.elsevier.com/locate/jslw

The effects of working memory and L2 proficiency on L2 writing

Olena Vasylets [a,*], Javier Marín [b]

[a] University of Barcelona, Spain
[b] University of Murcia, Spain

ARTICLE INFO

Keywords:
Working memory
L2 proficiency
L2 written production
Individual differences
Accuracy
Complexity

ABSTRACT

This study investigated whether the effects of working memory on second language (L2) writing performance depend upon the level of L2 proficiency. Fifty-six learners of L2 English at different levels of proficiency performed a complex working memory task in their native language (Spanish/Catalan), a standardized L2 proficiency test, and a narrative L2 writing task. L2 writing performance was assessed by means of the quantitative measures of accuracy, complexity, and fluency, and by means of holistic ratings of writing quality. The results showed that at low levels of proficiency, working memory was associated with higher accuracy, while at high levels of proficiency, there was a positive link between working memory and lexical sophistication. Our findings demonstrate a differential and nuanced involvement of working memory on L2 writing performance.

1. Introduction

Even in the native language, writing has been traditionally viewed as a skill acquired with effort and one "that is developed to immensely higher levels in some people than in others" (Bereiter & Scardamalia, 1987, p. 4). Second language (L2) writing, which can be even more constrained and more difficult (Weigle, 2005), is also characterized by high variability in the attained levels of proficiency. Because of the complexity of text composition, writing proficiency requires multiple skills and knowledge resources, including linguistic knowledge, knowledge of genre and rhetorical requirements, metacognitive knowledge of the writing process, and fluency in retrieval of linguistic knowledge (Schoonen, Van Gelderen, Stoel, Hulstijn, & De Glopper, 2011). Because of its multicomponential nature, variability in writing proficiency can be attributed to numerous factors, including the capacity and efficiency of a writer`s cognitive resources (Olive, 2012). Working memory (WM) has always stood out as one of the central cognitive variables underpinning writing performance (Hayes, 2012; Kellogg, 1996). Recent advances in the conceptualization and assessment of WM (Wen et al., 2015), together with the evidence that WM can be trainable (Tsai, Au, & Jaeggi, 2016), have revitalized research on this cognitive trait in the area of second language acquisition (SLA) (Grañena, Jackson, & Yilmaz, 2016). Although the role of WM is well-justified theoretically (Hayes, 2012; Kellogg, 1996; Kellogg, Whiteford, Turner, Cahill, & Mertens, 2013) and has been proven empirically in L1 writing (Vanderberg & Swanson, 2007), the effects of WM in L2 writing are less clear (Ahmaddian & Vasylets, 2021; Kormos, 2012). Recent theoretical thinking (Baddeley, 2015) and empirical research (Serafini & Sanz, 2016) has suggested, however, that the relationship between WM and L2 performance may depend on other factors, such as the level of L2 proficiency. These new research developments are suggestive enough to hypothesize that the role of WM in L2 writing may also be moderated by L2 proficiency. The

* Corresponding author at: University of Barcelona, Department of Modern Languages and Literatures and English Studies, Gran Vía de les Corts Catalanes, 585, 08006, Barcelona, Spain.
E-mail address: vasylets@ub.edu (O. Vasylets).

https://doi.org/10.1016/j.jslw.2020.100786
Received 7 October 2019; Received in revised form 1 December 2020; Accepted 3 December 2020
Available online 29 December 2020
1060-3743/

few studies which tested this assumption (Kormos & Sáfár, 2008; Lu, 2015) have produced contradictory findings, and further investigation of this issue is required. In this present study, we therefore explore: (a) if WM relates to the quality of L2 performance, and (b) if the potential relationship between WM and L2 writing performance is moderated by the level of L2 proficiency.

2. Literature review

2.1. Working memory in L2 acquisition

The term *working memory* refers to a limited-capacity cognitive system involved in the manipulation and maintenance in active attention of the task-relevant information and the inhibition of task-irrelevant information (Baddeley, 2003). According to the model by Baddeley and Hitch (1974), WM consists of a central executive, which controls two modality-specific subsystems: (1) phonological loop, responsible for manipulation and retention of phonological information; and (2) visuo-spatial sketchpad, which works with visual and spatial information. Although WM has been long viewed as a fixed trait with a strict limit (Cowan, 2005), there is some evidence to suggest that its capacity can be expanded through targeted training (Schwaighofer, Fischer, & Bühner, 2015), with the potential beneficial effects on academic performance in general and on L2 learning in particular (Tsai et al., 2016).

Today, WM constitutes an influential construct in neuroscience (Chai, Abd Hamid, & Abdullah, 2018), cognitive and educational psychology (Gathercole & Alloway, 2008) and SLA (Grañena et al., 2016; Wen et al., 2015). The importance of WM derives, at least partially, from its attributed role in learning and language processing (Baddeley, 2003). In L1 studies, individual differences in WM have been related to the success in speech development (Adams & Gathercole, 1995), language comprehension (Daneman & Merikle, 1996), or effective multitasking during the performance of academic tasks (Pollard & Courage, 2017). Inspired by the positive results from L1 academic learning, many researchers have attributed an important role to WM in SLA, to the point of recognizing it as a core component of language aptitude (Skehan, 2002; Wen, Biedroń, & Skehan, 2017). Empirically, a positive role of WM has been demonstrated in various L2 domains, such as the ability to learn from oral feedback (Mackey & Sachs, 2012) or vocabulary and grammar learning (Martin & Ellis, 2012; Serafini & Sanz, 2016) (for a review, see Juffs & Harrington, 2011; Williams, 2012). The most recent meta-analysis by Linck, Osthus, Koeth, and Bunting (2014) has reported a positive association between working memory capacity (WMC) and L2 outcomes, with an estimated population effect size (p) of .255. However, this positive influence does not manifest itself in an across-the-board manner, as numerous SLA studies have also found null WM effects (e.g., Grey, Cox, Serafini, & Sanz, 2015). Recently, a number of researchers (Linck et al., 2014; Serafini & Sanz, 2016) have suggested that the level of L2 proficiency can moderate the impact of WM on L2 development and performance.

2.2. The moderating role of L2 proficiency in the relationship between working memory and L2 processes and products

As pointed out by Williams (2015), "the relationship between WMC and L2 processing and learning is far more complex and nuanced than originally envisaged" (p. 301). The complexity of the overall picture of the role of WM in SLA could be due, at least in part, to the moderating influences of other factors, such as the level of L2 proficiency (Wen et al., 2015). From a theoretical standpoint, the moderating role of L2 proficiency could be explained through the links between WM and long-term memory (LTM). Many influential theorists have defined WM as an integral part of LTM (Baddeley, 2003; Cowan, 2005). Similarly, in the integrated WM-SLA framework by Wen et al. (2015), "the putative components of WM are all posited to be interacting bidirectionally with LTM … inhabited by learners' L1 mental lexicon and grammar as well as their L2 knowledge/proficiency" (p. 52). This posited two-way connection between WM and L2 knowledge allows the hypothesis that L2 proficiency may influence the mental operations in WM during the performance of SLA-relevant tasks. The idea that the effects of cognitive ability on L2 outcomes may be contingent on the level of L2 proficiency is not new and has its genesis in the studies of language aptitude (Robinson, 2005). For example, the meta-analysis by Li (2015) showed that the role of language aptitude in L2 outcomes was more critical at lower levels of L2 proficiency.

These SLA findings and theoretical contentions resonate with ideas in cognitive psychology, which contemplates different scenarios for the interactive effects of WM and knowledge on cognitive performance. For example, some theorists (e.g., Ackerman, 1988) adhere to the idea of compensation, which views WM as a mechanism that can compensate for the lack of knowledge and, thus, help maintain an optimal level of performance. Hence, a more prominent role of WM in task performance is attributed at lower levels of knowledge. An alternative scenario has been proposed by the *rich-get-richer* hypothesis (Hambrick & Engle, 2002). In this view, WM is seen as a conduit of knowledge during performance. In other words, WM is seen as a facilitating mechanism that allows individuals with higher WMC to benefit from their knowledge to a greater extent as compared to individuals with lower WMC. This theoretical conception predicts that variation in WMC would contribute to performance differences specifically at high levels of knowledge.

These theoretical ideas from cognitive psychology are suggestive of the possibility of different scenarios for the moderating effects of L2 proficiency on the links between WM and L2 outcomes. The mixed nature of the available evidence confirms this supposition. For example, Serafini and Sanz (2016) found that WM facilitated morphosyntactic development only for lower level learners (see also Dussias & Piñar, 2010; Sagarra & Herschensohn, 2010). Other studies, however, reported a completely opposite pattern of findings. Thus, the positive correlation between WMC and L2 performance was found to be restricted only to high proficiency learners in L2 oral production (Gilabert & Muñoz, 2010) or reading comprehension (Joh & Plakans, 2017). These findings provide a starting basis for our own empirical investigation, which aims to examine the potential contribution of WM to L2 writing through the prism of the potential interaction of WM with the level of L2 proficiency.

2.3. The role of working memory in L1 and L2 writing

In the theoretical models of writing, WM has been posited to be a central cognitive resource need for composition (Hayes, 2012; Kellogg, 1996). According to Kellogg's (1996) L1 writing model, for example, the central executive is implicated in all higher-level writing processes, including planning, linguistic encoding, and revision. Importantly, an acknowledged limitation of the original 1996 model is "its failure to consider the role of domain expertise" (Kellogg et al., 2013, p. 167). In line with the compensation theory from cognitive psychology (Ackerman, 1988), Kellogg et al. (2013) theorized that WM could be less taxed for writers with higher degrees of domain-specific (and/or linguistic) knowledge (see also McCutchen, 2011). This theoretical supposition has support from various empirical studies, which found that the composition of writers with more efficient WM resources displayed higher quality of L1 writing performance (Hoskyn & Swanson, 2003; Vanderberg & Swanson, 2007). As compared to L1, composing in L2 may entail additional difficulties, which can stem from the gaps in L2 knowledge and/or from insufficiently automatized access to L2 linguistic representations (Weigle, 2005). Consequently, it would be feasible to hypothesize that WM could play even a more important role for L2 writers. This supposition, however, is not well supported by the available research findings, which are mixed and inconclusive.

For instance, an early study by Adams and Guillot (2008) explored a relationship between WMC and writing of French/English bilinguals. The participants performed a timed picture description task in English and in French, which was assessed on a scale of zero to 10 by four raters. A positive link was found between L2 English writing and the phonological component of WM. A later study by Bergsleithner (2010) reported a positive relationship between complex WM, accuracy, and subordination of L2 writing description tasks performed by adult L1 Brazilian learners of English. A positive link between complex WM and some aspects of accuracy was also found by Zalbidea (2017), who examined the written argumentative performance of L2 Spanish learners. However, Zabihi (2018) reported that WMC was positively related to fluency (length of T-unit) and subordination, but negatively affected accuracy (ratio of error-free T-units) of L2 written narratives performed by adult Persian learners of L2 English. More recently, Michel, Kormos, Brunfaut, and Ratajczak (2019) investigated the role of WM in the performance of young Hungarian learners of L2 English. The participants performed a series of WM tasks and three different types of writing tasks, which were rated on a scale of zero to four based on the TOEFL Junior Comprehensive performance descriptors. Except for the academic editing task, there were no significant associations between WM functioning and L2 writing tasks. Finally, Mavrou (2020) reported a positive link between the updating function of WM and accuracy and subordination of L2 written video-retellings performed by the adult learners of L2 Spanish.

Of specific relevance to our study are the investigations which took into account the level of L2 proficiency when examining WM effects in L2 writing. The study by Kormos and Sáfár (2008) involved beginning and pre-intermediate secondary school learners in Hungary. The participants performed a phonological short-term memory (PSTM) test, a complex WM test (taken only by the beginning learners), and three writing tasks of different genres. The scores from the Cambridge First Certificate Exam were used as a measure of proficiency. Written texts were assessed for content and accuracy by two raters. A moderate correlation between PSTM test and L2 writing was found for pre-intermediate learners, but not for beginners. The authors concluded that PSTM may play a different role in the case of beginners and pre-intermediate students. Different results were obtained in the study by Lu (2015), in which Chinese learners of English wrote an argumentative essay in L2 English under a timed condition and performed a complex WM test (operation span) in L1 Chinese and in L2 English. The level of L2 knowledge was assessed by means of productive and receptive L2 vocabulary knowledge tests. Written performance was assessed by two raters in terms of content and organization on a scale of one to 15. The analysis showed absence of correlations between the measures of WMC and L2 written performance.

In sum, although some studies have obtained evidence for the positive role of WM in L2 written performance (e.g., Bergsleithner, 2010), other investigations have obtained results which were mixed (Zabihi, 2018) or practically null (Michel et al., 2019). Such a mixed nature of the overall findings makes it challenging to arrive to any firm conclusion concerning the role of WM in L2 writing. It is also unclear if proficiency moderates the links between WM and L2 writing performance. The available investigations on this matter are very scarce and have certain methodological limitations. For example, in the study by Kormos and Sáfár (2008), the more advanced learners took only the PSTM, which limits the study's conclusions concerning the role of complex WM in L2 writing. On the other hand, Lu (2015) divided the participants into high and low proficiency groups on the basis of their vocabulary scores, which might be questionable. In sum, more investigations are required to determine if and how the level of L2 proficiency may impact the involvement of WM in L2 writing.

It is also challenging to compare results across available studies because of the differences in the way the quality of L2 writing was assessed. While some studies employed holistic ratings, which give a measure of overall writing quality (e.g., Kormos & Sáfár, 2008), others have relied on quantitative CAF measures, which tap into the discrete areas of L2 performance (Zabihi, 2018). There are certain limitations, however, in the CAF measurement practices of previous studies, specifically in the assessment of L2 complexity. As argued by Bulté and Housen (2012), L2 complexity of monologic production minimally consists of linguistic (syntactic and lexical) complexity and propositional complexity. However, the assessment of L2 performance in previous studies does not fully reflect the multidimensional nature of L2 complexity. Concerning syntactic complexity, for example, the available investigations have overwhelmingly focused on subordination, leaving out other important syntactic subdimensions such as coordination or nominal complexity. As for lexical complexity, only Zalbidea (2017) looked into lexical diversity, while the role of WM in lexical sophistication of L2 writing has not been yet explored. Finally, none of the previous studies assessed propositional complexity, which refers to the number of information or idea units encoded in the message (Bulté & Housen, 2012). Although a newcomer to SLA research, idea units have proved to be instrumental in shedding light on such important issues as the longitudinal development of L2 English proficiency (Larsen-Freeman, 2006), the effects of task complexity on L2 oral and written production (Vasylets, Gilabert, & Manchón, 2017), differences in L2 complexity in speech and writing (Vasylets, Gilabert, & Manchón, 2019) or the way L2 learners achieve communicative success in L2 oral versus written tasks (Vasylets, Gilabert, & Manchón, 2020).

Thus, in light of the identified gaps, this study aims to contribute to the knowledge about the influence of WM on L2 written performance. We further explore if the potential link between WM and L2 writing is moderated by the level of L2 proficiency. To enhance comparability with previous research, we employ both holistic rating and CAF measures to assess L2 written performance. We also extend the range of L2 writing measures by incorporating the measures of propositional complexity (idea units), coordination, nominal complexity, and lexical sophistication.

3. The present study: research questions

The following research questions were posed in this study:

RQ1: Is there a significant relationship between WMC and L2 writing performance, as assessed by means of holistic ratings and by means of the quantitative measures of accuracy, complexity (linguistic and propositional), and fluency?

RQ2: If so, does L2 proficiency moderate the relationship between WMC and L2 writing performance, as assessed by means of holistic ratings and by means of the quantitative measures of accuracy, complexity (linguistic and propositional), and fluency?

Solid theoretical justification (Hayes, 2012; Kellogg, 1996; Kellogg et al., 2013) coupled with empirical evidence (albeit tentative because of its mixed nature) make it feasible to hypothesize a positive link between WMC and L2 written performance, as assessed by both quantitative CAF measures and holistic ratings. Provided the existence of conflicting theoretical predictions (Hambrick & Engle, 2002) and the mixed nature of empirical findings (Kormos & Sáfár, 2008; Lu, 2015), a non-directional hypothesis is advanced concerning the moderating role of L2 proficiency. Thus, we hypothesize that L2 proficiency would moderate the effects of WMC on L2 written performance.

4. Method

4.1. Participants

The participants were 59 native Spanish/Catalan learners of L2 English (12 male, 47 female; age: $M = 19$ years old, $SD = 8$). The participants were second year university students at a public Spanish university, and they all came from the same class. Spanish and Catalan were dominant L1s for 32 and 24 participants respectively, while three of them declared themselves to be balanced Spanish/Catalan bilinguals. All participants had received similar L1 and L2 writing instruction in accordance with the public educational program in Spain, and they had been exposed to similar methods of English instruction, which represented a combination of communicative and form-focused language teaching approaches. On average, the participants had received 12.3 years of English instruction prior to the study ($SD = 3.08$), and they displayed a range of L2 proficiency levels, with $n = 7$ learners belonging to B1, $n = 21$ to B2, $n = 21$ to C1, and $n = 7$ to C2 level as according to the Common European Framework of Reference (CEFR) (Council of Europe, Council for Cultural Cooperation, Education Committee, & Modern Languages Division, 2001). As a reward for participating in the study, the participants were awarded credit points in one of their classes.

4.2. Instruments

4.2.1. Measure of WM capacity

To evaluate WMC, we employed a complex verbal span task, which is a Spanish/Catalan version of Unsworth, Heitz, Schrock, and Engle's (2005) original test (Gilabert & Muñoz, 2010). The test represents a computerized reading span, which requires the participants to make judgments of the semantic plausibility of individual sentences while trying to remember a set of unrelated letters. The test proceeds as follows: individual sentences are presented on the screen in written form and the participants are required to click either a *true* or *false* box, depending if the sentence is correct (true) or if it contains a semantic deviation (false). Next, the participants see a letter to be remembered. Then, the procedure repeats. In the recall phase, the participants have to click the box next to the appropriate letter in the same order in which the letters were presented. The size of sentence-letter string ranges from three to seven. In total, the test contains 80 sentences, half of which are correct and half of which are semantically deviant. Sentence length ranges from 10 to 15 words. The sentences and letters are presented at the rate of approximately one time every 750 milliseconds, and the whole test takes between 20 and 30 min to complete. The test generates two scores of WMC: the absolute score and the partial-credit score. The absolute score is more stringent, as one point is given only to the set in which all the items were correctly recalled. The partial-credit unit scoring assigns proportional scores depending on the number of the correctly recalled items. In line with previous research (Conway et al., 2005), absolute and partial unit scores were highly correlated ($r = .85$, $p \leq .001$), and an absolute score was employed for the purposes of this study.

4.2.2. Measure of L2 proficiency

To assess the level of L2 linguistic knowledge, the pen-and-paper version of the Oxford Quick Placement Test (QPT) (UCLES, 2001) was employed. QPT consists of the subtests on different language skills, focusing primarily on vocabulary and grammar knowledge. The test contains 60 items and requires learners to select the appropriate answer to three- or four-option multiple-choice items targeting a range of grammatical structures in English, such as prepositions, tense, aspect, gender, and number agreement. The test also contains a number of exercises which tap into knowledge of pragmatics and vocabulary. The vocabulary exercises represent fill-in-the-gap exercises which require learners to choose, out of the provided options, the word (all options are nouns) or phrase

(lexical collocations) which best fits the spaces in the text. QPT takes up to 30 min to complete. The scoring system of the test allows for establishing an equivalence between the test scores and the levels of CEFR.

4.2.3. Writing task

To obtain the L2 writing sample, a video-retelling task deemed to elicit narrative discourse was employed. A narrative task was chosen, as this type of discourse is considered to be one of the major and most universal types of discourse in everyday language production (Berman & Slobin, 2013), hence the importance of the development of narrative skills in an L2. In terms of its prototypical features, narratives have a clearly defined macrostructure, which evolves around temporal and causal relations. Narratives typically focus on actors (e.g., people, animals) and their actions and motivations, and consist of a succession of events in chronological order (Berman & Slobin, 2013). As a stimulus to elicit a sample of narrative discourse, an episode from the "Simon's Cat" series was employed. "Simon's Cat" is an animated series featuring adventures of a cat and his owner Simon. The series consists of episodes that are two to five minutes long with a self-contained plot. The episodes have a musical background, but the characters do not speak and no other linguistic input (e.g., captions) is provided. The task instructions asked the participants to watch the episode and retell in writing what they had seen. There was a requirement of a minimum text length of 200 words. The task was performed on the computer, but no access to dictionaries or any other external sources was available. The participants were prompted to start writing immediately after having watched the clip, and no specific instructions were given regarding planning or revision. There was no specific time restriction for the writing task performance.

4.2.4. Piloting

Prior to the experiment, the WM test and the writing task were piloted in the language lab with a group of 10 learners with a similar profile to the participants of the study. The learners were provided with the instructions and performed the tasks individually on the computer. All learners first started with the WM test and then proceeded with the writing task. The piloting showed that the instructions of both WM test and writing task were comprehensive. The writing task appeared to be engaging and elicited narrative discourse, as expected. During the piloting, the learners finished the WM test within a range of 20−35 min and the writing task within a range of 15−30 min. No piloting of QPT was undertaken because it represents a standardized test.

4.3. Procedure

The tests and the writing task were completed in the following order:

1. Background questionnaire
2. Quick Oxford Placement Test (QPT)
3. Working memory test
4. Writing task

Data collection took place during two sessions in a computer lab. During the first session, the participants completed a background questionnaire and the paper version of the QPT, which took between 20 and 30 min to complete. During the second session, the participants performed the WM test and the writing task on individual computers. The participants started with the WM test, with all the instructions appearing on the screen. Depending on their self-perceived L1 dominance, the participants could choose between the Spanish or Catalan versions of the test. It took between 20 and 30 min to complete the WM test. After completing it, the participants proceeded with the writing task. The majority of the participants finished the writing task within 30 min.

4.4. Analysis of L2 written production

4.4.1. CAF measures

We strived to choose the measures which would ensure the multidimensional analysis of performance while simultaneously avoiding redundancy (Norris & Ortega, 2009). For lexical complexity, we employed the measures of lexical diversity and sophistication. D_Tools software (Meara & Miralpeix, 2006) was used to obtain D-value (Malvern, Richards, Chipere, & Durán, 2004), which was used as an index of lexical diversity. This index provides an indication of the degree of word's repetition in a text. Program RANGE (Heatley, Nation, & Coxhead, 2002) was used to calculate the Advanced Guiraud index. This measure indicates the extent to which a learner employs "advanced" words in production. As Advanced Guiraud captures only those words which do not figure among the most frequent words, this measure can be taken as an index of lexical sophistication (Bulté & Housen, 2014).

To tap into syntactic complexity, we assessed coordination, subordination, and nominal complexity. For complexity via coordination, the ratio of coordinate phrases (coordinate phrases/total number of clauses) was obtained. To gauge subordination, we calculated the ratio of dependent clauses (dependent clauses/total number of clauses). The ratio of complex nominal structures (complex nominals/total number of clauses) was a measure of nominal complexity. Complex nominals included: (a) nouns plus adjective, possessive, prepositional phrase, adjective clause, participle, or appositive; (b) nominal clauses; and (c) gerunds and infinitives in subject position. All syntactic measures were obtained with the help of the Synlex software (Lu, 2010).

The analysis of propositional complexity was performed manually. As a first step in the analysis, each text was segmented into idea units using guidelines from our previous research (Vasylets, 2017). In these guidelines, an idea unit represents a semantic unit of discourse, which is defined as a meaningful, semantically integral chunk of discourse (Chafe, 1994). The guidelines define a

"meaningful" chunk of discourse as linguistic material which makes sense, cognitively, to the writer who produces the idea and to the reader who interprets the idea. "Meaningful" also implies that an idea conveys information about an event, state, referent, or location in space or time. "Semantically integral" means that an idea conveys a message which constitutes an undividable block of information. For higher validity and reliability, the guidelines complement the primary semantic criteria with the syntactic criteria. Syntactically, an idea unit typically (but not exclusively) represents a clause. The segments of discourse bound by coordinate relationship represent separate ideas. In the case of subordinate constructions, the main and subordinate elements represent independent ideas in the case of loose conceptual connection, and form part of the same idea in the case of strong conceptual linkage, as in the case of restrictive relative clauses, clauses introduced by "that" or "to," or adverbial clauses in the final position. A single idea unit constituted by the main and subordinate clause/s are labelled as extended ideas (Vasylets, 2017). Provided that extended ideas contain, by default, more than one semantically meaningful constituent, this type of idea is considered as informationally dense. To assess propositional complexity, we calculated the total number of idea units and the ratio of extended idea units (extended idea units/all idea units).

To assess accuracy, the measure of the number of errors per 100 words (all errors/words x 100) was employed. We took into account errors in grammar and vocabulary. Spelling and punctuation errors were not counted. The criteria to define an error were set against Standard English. The mean length of T-unit (T-units/tokens) was calculated to assess writing fluency (Wolfe-Quintero, Inagaki, & Kim, 1998; Zabihi, 2018).

4.4.2. Holistic rating

Six university teachers, who had ample experience with the use of rating rubrics, were recruited to perform the holistic assessment of the texts. A pseudo-random procedure was employed to distribute the texts among the raters in such a way that: (a) all raters had to assess a similar number of texts; and (b) each text was assessed by at least three raters. As a guiding rubric, The Independent Writing Rubrics from the TOEFL iBT Test was employed.

This instrument assesses writing quality on a scale of zero to five based on the topic development, text structure, coherence, syntactic variety, vocabulary range, idiomatic language, and errors. After participating in the training session, the raters assessed the written texts on their own time. To obtain the final holistic score for each writing sample, the scores from different raters were averaged.

4.4.3. Intra-rater reliability

To assess intra-rater reliability for the quantitative measures, one of the researchers of the study recalculated 30% of data for idea units, extended idea units, and errors in one month after the first analysis. Cohen's kappa was of 0.9, 0.88 and 0.86 for idea units, extended ideas, and accuracy, respectively. To assess the reliability of the holistic ratings, Cohen's kappa coefficients were calculated for each pair of raters who rated at least eight of the same texts; the overall mean of the Cohen's kappa coefficients for the holistic ratings was 0.85. All Cohen's kappa coefficients obtained for quantitative and holistic measures were within the acceptable range of reliability (Cohen, 1988).

4.5. Statistical analysis

For the statistical analysis, descriptive statistics were computed for the WM test, L2 proficiency test, and outcome measures. Next, we ran a correlation between the outcome measures, and we also ran a correlation to probe into relationship between WMC and the measures of L2 writing performance. To answer the question about the mechanisms by which WM transmits its potential effects to L2 writing, we performed a moderation analysis, which represents a regression-based path analysis (Field, 2009). We deemed this analysis appropriate, as it allows for exploring the association between the predictor variable (WMC in this case) and outcomes (CAF and holistic rating), and if this potential association is moderated: i.e., whether its size, sign, or strength depends on or can be predicted by a third variable, which is a moderator (L2 proficiency). An important advantage of this analysis is that both WMC and L2 proficiency were treated as continuous variables, which allowed for an understanding of the full range of the participants' performance (see also Murphy & Creux, 2020). To perform moderation analysis, we employed a SPSS PROCESS macro, which represents a computational tool for path analysis-based moderation (Hayes, 2018). This versatile modelling tool has been widely employed in the empirical literature of many disciplines, including writing research (Baaijen, Galbraith, & de Glopper, 2014). Specifically, we employed PROCESS Model 1 to produce the output to probe WMC and L2 proficiency interactive effects on the measures of L2 written production. Separate analysis was performed for each of the CAF dimensions and for holistic rating (Field, 2009). Fig. 1 summarizes the relationship tested in the current study.

Before performing parametric tests, the data were checked for compliance with the assumptions of normal distribution,

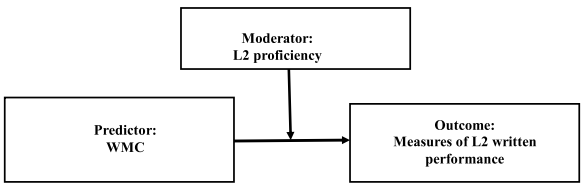

Fig. 1. Overview of the Hypothesized Moderation Model.

multicollinearity, and homoscedasticity, with no serious violations noted. Analyses were repeated with and without the most extreme cases, and relevant coefficients remained significant. The analysis showed that the predictor (WMC) and the moderator (proficiency) were not highly correlated: $r = .21$, $p = .11$.

5. Results

Table 1 provides descriptive statistics for the WM test and L2 proficiency test. Minimum (26) and maximum (63) values indicate that the range in the levels of WMC was rather considerable. Minimum (34) and maximum (58) values for the QPT show the level of L2 proficiency ranged from B1 to C2 levels. Table 2 reports descriptive statistics on the L2 proficiency scores as divided by the corresponding CEFR levels, showing that the participants belonged, in their majority, to the upper-intermediate and advanced L2 proficiency levels.

Following Norris and Ortega's (2009) recommendation to avoid redundancy in the measurement practices, we calculated Pearson correlation coefficients to check for interdependency between outcome measures (see Table 3). This analysis revealed that holistic ratings correlated negatively with the number of errors ($r = -.492$, $p \leq .05$) and positively with idea units ($r = .389$, $p \leq .01$), lexical diversity ($r = .384$, $p \leq .01$), and lexical sophistication ($r = .557$, $p \leq .01$). These results showed that during the holistic assessment of the L2 texts, the judges relied more on lexical complexity, accuracy, and semantic content, rather than on syntax and grammar (for similar results, see Kuiken, Vedder, & Gilabert, 2010). The majority of significant positive correlations were obtained for the measures of lexical complexity. Thus, apart from the positive relationship with the holistic rating, D-value correlated positively with the number of idea units ($r = .319$, $p \leq .05$), subordination ($r = .330$, $p \leq .05$), and nominal complexity ($r = .369$, $p \leq .05$), while Advanced Guiraud also correlated with idea units ($r = .429$, $p \leq .01$) and nominal complexity ($r = .280$, $p \leq .05$). Also, the measure of coordination positively correlated with nominal complexity ($r = .312$, $p \leq .05$) and with the length of T-units ($r = .386$, $p \leq .05$). These connections provide evidence of the interdependence of the CAF traits highlighted in previous research (see contributions to Housen, Kuiken, & Vedder, 2012; Norris & Ortega, 2009). The highest correlation ($r = .568$, $p \leq .01$) was obtained between D-value and Advanced Guiraud, which shows that the measure of lexical diversity and sophistication tap into two highly related constructs, which share for about 32% of variance. Finally, the fact that idea units correlated with the measures of lexical (but not syntactic) complexity provides evidence (albeit indirect) that the construct of idea unit constitutes primarily a semantic unit, rather than a syntactic unit of discourse.

To answer the main questions of this study, we first calculated descriptive statistics for the CAF measures and holistic ratings, as well as their correlations with WMC (Table 4). The small negative correlation between WMC and the number of errors ($r = -.28$, $p \leq .05$) indicates that higher WMC is associated with higher writing accuracy.

To explore further the relationship between WMC and the measures of L2 written performance, we performed a series of moderation analyses with WMC as a predictor variable, the level of L2 proficiency as a moderator variable, and the CAF measures and holistic ratings as the outcomes (i.e., dependent variables). To perform the analyses, Model 1 from the SPSS macro PROCESS (Hayes, 2018) was used. Z-scores of the predictor, moderator, and outcomes variables were employed for the moderation analysis. The results show that the overall model (see Table 5) was significant for the Advanced Guiraud, which measures lexical sophistication ($F(3,52) = 4.66$, $p \leq .001$, $R^2 = .21$), and for the number of errors ($F(3,52) = 22.6$, $p \leq .001$, $R^2 = .56$). These results show that WM and L2 proficiency explained more than half of the total variance in L2 accuracy. And for lexical sophistication, about one fifth of the total variance was on account of WM and L2 proficiency.

As shown in Table 6, the interaction between WMC and L2 proficiency is significant for the Advanced Guiraud ($b = .29$, $SE = .14$, $t(52) = 2.10$, $p = .04$, 95% $CI = [0.04\ 0.57]$) and for the number of errors ($b = .23$, $SE = .10$, $t(52) = 2.19$, $p = .03$, 95% $CI = [0.02\ 0.43]$).

Simple slope analysis further showed that the relationship between WMC and lexical sophistication was significantly positive at higher levels of proficiency ($b = .47$, $SE = .19$, $t(52) = 2.49$, $p \leq .001$, 95% $CI = [0.09\ 0.86]$), but was not significant at lower levels of proficiency ($b = -.17$, $SE = .20$, $t(52) = -.82$, $p = .24$, 95% $CI = [-0.58\ 0.24]$). On the other hand, the relationship between WMC and errors was significantly negative at lower levels of proficiency ($b = -.42$, $SE = .15$, $t(52) = -2.79$, $p = .01$, 95% $CI = [-0.73\ -0.11]$), but was not significant at higher levels of proficiency ($b = .07$, $SE = .14$, $t(52) = .53$, $p = .56$, 95% $CI = [-0.21\ 0.35]$).

6. Discussion

The aim of this study was to investigate the potential effects of WMC on L2 writing performance. To gain a more nuanced understanding of this phenomenon, the impact of WMC was explored through the prism of its potential interaction with L2 proficiency, and we employed both holistic ratings and discrete CAF measures to assess the quality of L2 writing.

Our initial prediction was that WMC would be positively related to the measures of L2 writing quality. We also expected that L2

Table 1
Descriptive Statistics for WM and L2 roficiency Tests.

Variables	Mean	SD	Median	Min	Max
Working memory	39	11.2	36.5	26	63
L2 proficiency	47	6.32	47.5	34	58
$N = 56$					

Table 2
Descriptive Statistics for the Correspondence of L2 Proficiency Scores to the CEFR Levels.

CEFR levels (OPT scores)	N	Mean	SD	Median	Min	Max	Percent
B1 (30–39)	7	37.57	1.81	38	34	39	12.5
B2 (40–47)	21	42.95	2.69	42	40	47	37.5
C1 (48–54)	21	51	2	51	48	54	12.5
C2 (55–60)	7	56.57	1.34	57	55	58	37.5
Total	56						100

Table 3
Pearson Correlations Between the CAF Measures and Holistic Rating of L2 Written Texts.

Variables	D-value	Adv. Guiraud	Ratio coord. phrases	Ratio dep. clauses	Ratio compl. nom.	Num. idea units	Ratio exten. ideas	Errors per 100 words	Mean length T-unit	Holistic rating
D-value	1									
Advanced Guiraud	.568*	1								
Ratio of coordinate phrases	.129	.110	1							
Ratio of dependent clauses	.330*	.004	.208	1						
Ratio of complex nominals	.369**	.280*	.312*	.183	1					
Number of idea units	.319*	.429**	.141	−.071	.083	1				
Ratio of extended ideas	−.042	−.131	.022	.229	−.079	−.138	1			
Errors per 100 words	−.066	−.311	.071	.240	.152	−.167	.056	1		
Mean length of T-unit	−.080	−.107	.386*	.008	.173	.070	.173	.157	1	
Holistic rating	.384**	.557**	.223	−.016	.216	.389**	.085	−.492*	.006	1

Note. $N = 56$; *$p \leq 0.05$; **$p \leq 0.01$.

Table 4
Descriptive Statistics for the Measures of L2 Writing Performance and Correlations with WMC Score.

Performance measures	Mean	SD	Min	Max	Correlations with WMC	Sig. (2-tailed)
Lexical complexity						
D-value	41.68	9.24	23.10	69.00	.06	.57
Advanced Guiraud	1.38	.40	.5	2.31	.23	.09
Syntactic complexity						
Ratio of coordinate phrases	.24	.13	.00	0.60	.21	.12
Ratio of dependent clauses	.35	.11	.14	.64	−.03	.85
Ratio of complex nominals	.65	.19	.33	1.13	.14	.31
Propositional complexity						
Number of idea units	23.91	8.18	9	47	.13	.35
Ratio of extended ideas	.25	.10	.06	.50	−.22	.10
Accuracy						
Errors per 100 words	4.22	2.19	.99	9.64	*-.28	*.04
Fluency						
Mean length of T-unit	8.47	1.92	1.9	11.29	−.18	.26
Holistic rating	3.48	.91	3.83	1	.16	.21

Note. $N = 56$; *$p \leq 0.05$.

proficiency would moderate the potential impact of WM on L2 writing. Our results provided partial support to the prediction about the effects of WMC, as we found that WMC contributed positively only to the selected dimensions of L2 writing performance. Specifically, we found a positive link between WMC and the dimensions of accuracy and lexical sophistication. Concerning the moderation effects of

Table 5
Predicting Measures of L2 Written Performance from WMC Scores, L2 Proficiency and Their Interactions.

Performance measures	F (3,52)	p	R^2
Lexical complexity			
D-value	2.46	.07	.12
Advanced Guiraud	4.66	**.00	.21
Syntactic complexity			
Ratio of coordinate phrases	1.16	.33	.06
Ratio of dependent clauses	.06	.98	.00
Ratio of complex nominals	1.08	.36	.05
Propositional complexity			
Number of idea units	1.22	.31	.06
Ratio of extended ideas	1.55	.21	.08
Accuracy			
Errors per 100 words	22.6	**.00	.56
Fluency			
Mean length of T-unit	1.27	.29	.07
Holistic rating	1.26	.29	.06

Note. N = 56; **p ≤ 0.001.

Table 6
Results for the WMC and L2 Proficiency Interaction Effects for the Measures of L2 Written Performance.

					95% CI	
Performance measures	b	SE	t	p	LL	UL
Lexical complexity						
D-value	.28	.14	1.95	.06	−.00	.57
Advanced Guiraud	.29	.14	2.10	.04*	.04	.57
Syntactic complexity						
Ratio of coordinate phrases	−.13	.15	−.89	.37	−.44	.17
Ratio of dependent clauses	−.01	.15	−.11	.91	−.33	.29
Ratio of complex nominals	.22	.15	1.14	.14	−.07	.52
Propositional complexity						
Number of idea units	−.13	.15	−.91	.36	−.44	.16
Ratio of extended ideas	−.19	.14	−1.29	.10	−.49	.10
Accuracy						
Errors per 100 words	.23	.10	2.19	.03*	.02	.43
Fluency						
Mean length of T-unit	−.24	.15	−1.58	.12	−.55	.06
Holistic rating	−.16	.14	−1.17	.24	−.45	.11

Note. $N = 56$; $*p \leq 0.05$.

L2 proficiency, a complex pattern of findings was obtained, as we found that the nature of the interactive effects of WMC and L2 proficiency depended upon the performance dimension.

Thus, we found that higher WMC was positively related to accuracy only for lower proficiency learners. Consistent with our results, a positive relationship between WMC and L2 writing accuracy was also obtained in Bergsleithner (2010) and Zalbidea (2017) for adult L2 writers. This result also aligns with previous SLA investigations which reported a positive link between WMC and morphosyntactic development for beginning and intermediate learners (Dussias & Piñar, 2010; Sagarra & Herschensohn, 2010; Serafini & Sanz, 2016). Our findings for accuracy are consistent with the theoretical contention advanced by Kellogg et al. (2013) and also with the theory of compensation (Ackerman, 1988), which predicted a greater contribution of WM to task performance at lower levels of knowledge. It could be suggested that writers with higher WMC would find themselves better equipped to compensate gaps in L2 proficiency, successfully resolving various linguistic challenges related to the ability to communicate without errors. There can be different ways in which WM can contribute to L2 writing accuracy. Firstly, accuracy could relate to the monitoring stage of L2 writing: higher WMC could enhance reading and revision processes, with a concomitant decrease in the errors in the written text. Higher WMC could also enhance searches for the linguistic alternatives and thus prevent errors stemming from gaps in L2 knowledge. Additionally, WM could be instrumental in supressing the competing erroneous linguistic representations and thus could support error-free production of

newly acquired and not yet stabilized L2 items (see also Wolfe-Quintero et al., 1998). Importantly, gaps of L2 knowledge and a lack of automaticity are specifically problematic at lower levels of proficiency, hence the enhanced role of WM for accuracy for less proficient L2 writers.

Another relevant finding was a link between WMC and lexical sophistication. Importantly, in the area of lexis, the interactive effects of WM and L2 proficiency worked in the completely opposite way as compared to accuracy. Thus, we found that WM enhanced lexical sophistication only for writers with higher L2 proficiency. This finding aligns with other studies (e.g., Gilabert & Muñoz, 2010; Joh & Plakans, 2017) which found that WMC was associated with better L2 performance only for high proficient learners. The rich-get-richer hypothesis (Hambrick & Engle, 2002) provides a coherent explanation for our findings for lexical sophistication. This theory posits that the efficiency of WM is an important contributor to task performance specifically at high levels of knowledge. As applied to L2 writing in particular, we could suggest that lexical complexity could be related to the higher-order writing processes, which all heavily draw on WM resources. Thus, higher efficiency in WM could facilitate the preparation of a complex conceptual plan, calling for more lexically sophisticated linguistic encoding. During linguistic formulation, WM can enhance linguistic searches, giving the conceptual plan its corresponding lexically complex form. During monitoring, higher WMC can be instrumental, inter alia, for detecting repetition or for revising/improving lexical choices, with a concomitant boost in lexical complexity. However, for these positive WM influences to take place, learners might need to reach a certain threshold in their level of L2 knowledge, hence the positive effects for lexical complexity only for higher proficient learners. We could suggest that WMC was not related to the lexical sophistication of low proficient L2 writers because their vocabulary was not sophisticated enough for WM to make a meaningful impact.

An issue which draws special attention is the difference in the findings for accuracy and lexical sophistication. Accuracy and lexical sophistication constitute distinct performance areas that compete for limited cognitive resources (Housen & Kuiken, 2009). Allocation of cognitive resources during L2 performance can depend upon various factors, such as the nature of the task (Skehan & Foster, 1997) or learners' natural predisposition towards complexity or accuracy (Ortega, 1999), or it can also be a strategic decision of learners (O'Malley & Chamot, 1990). The scenario of the strategic distribution of cognitive resources is posited as particularly feasible in such tasks as writing, which entails control of time and a self-paced manner of production (Vasylets et al., 2017). In relation to our findings, we could suggest that learners of lower proficiency may have strategically geared their cognitive resources towards goals which they perceived as affordable, such as focus on form and the associated accuracy of L2 writing. Importantly, the development of L2 proficiency can cause a shift in learners' priorities, inducing them to devote their cognitive capacity to new goals, such as the production of more complex language. The theoretical idea that L2 proficiency can influence learners' order of priorities in the allocation of WM resources, is, however, tentative and needs empirical testing.

Concerning other performance dimensions, there were no links between WM and propositional complexity, syntactic complexity, or fluency. The absence of the relationship with syntactic complexity was surprising and contradicts previous findings (Bergsleithner, 2010; Zabihi, 2018). Explanation for this finding could lie in the nature of our WM test in which the participants had to memorize letters and also judge if sentences contained semantic deviations. It could be tentatively suggested that if our WM test had included a grammaticality judgment task instead of the semantic judgments, different results would have been obtained concerning the link between WMC and L2 writing syntactic complexity (Alptekin, Erçetin, & Özemir, 2014). In contrast to Zabihi (2018), who found a positive connection between WMC and fluency, we obtained null results for this measure of performance. This discrepancy could be explained by the different writing conditions: while Zabihi (2018) gave 11 min for the learners to perform a task, there were no specific time restrictions (i.e., less time pressure) in our study. Finally, the absence of links between WMC and propositional complexity (idea units) could be attributed to the nature of the stimulus we employed to elicit the written performance. Thus, the task required participants to narrate some predetermined content, which was easy to remember. Because of the simplicity of the plot, WM might not have been substantially involved to render the ideational content, with the concomitant absence of links between WMC and propositional complexity. Further research should explore if the relationship between WMC and propositional complexity, as well as fluency, would remain the same in more creative and challenging writing tasks, such as argumentation or opinion essays.

Finally, we also found that there were no significant links between WMC and the holistic assessment of L2 writing performance. This finding has support in the literature, as virtually all previous L2 research reported the absence of connections between complex WM and holistic scores of L2 writing (Adams & Guillot, 2008; Kormos & Sáfár, 2008; Lu, 2015). Our findings align with other studies which also obtained different results with quantitative and holistic measures of L2 writing. Bulté and Housen (2014), for example, reported that holistic measures were more useful to detect L2 writing development over time. In our study, however, the CAF measures appeared to be more instrumental, which can be explained by the nature of the involvement of WM in L2 written quality. Our finding that WMC was linked only to accuracy and lexical sophistication attests a nuanced and selective involvement of this cognitive resource, which can play a role in some L2 writing dimensions and be less relevant in others. For this reason, the CAF measures, which tap into the discrete performance features, discerned WM effects better as compared to the holistic score, which integrates the assessment of various performance dimensions into one single score of overall L2 writing quality.

7. Conclusion and implications

Overall, we can conclude that in our study, higher WMC appeared as a positive asset in L2 writing. However, our results indicate the contribution of WM to L2 written performance might not always be independent or significant; rather, it could be influenced by the writers' L2 proficiency and it may depend upon the area of production. This study has a number of implications. Thus, our findings contribute to the growing evidence of the differential involvement of cognitive abilities in L2 processing and performance (Serafini & Sanz, 2016). In terms of the theoretical implications, our results confirm theories which contemplate a connection between WM and L2 knowledge (Wen et al., 2015). Also, the obtained findings can be useful for the language-learning potential of L2 writing theory

(Manchón & Vasylets, 2019): new knowledge about the role of cognitive individual differences can help refine the description of the mechanisms by which L2 writing advances general L2 proficiency. Our findings also provide evidence that the study of the relationship between WM and L2 writing can provide real leverage in the exploration of the dynamic change of learners' strategic priorities at different levels of L2 proficiency. And from a methodological standpoint, we showed the importance of using both quantitative and holistic measures in the assessment of L2 writing quality.

Knowledge about the role of WM in L2 writing may also have relevance for educational practices. As pointed out by Gathercole and Alloway (2008), failures of WM are often misclassified as a failure of motivation, with detrimental consequences for learning processes. For this reason, understanding that difficulties with L2 writing might also be a consequence of WM failures is important for the design of more effective tasks and more effective classroom management. Examples of tasks which could help learners avoid WM overload is providing learners with content support in the form of bullet points and macrostructure (Ong & Zhang, 2010), or providing a plot in the form of a cartoon (Kormos, 2011). Such provision of semantic and/or linguistic content could potentially alleviate the cognitive load posed by the task and allow for the gearing of the cognitive resources toward linguistic encoding and revision, with potential benefits for L2 writing quality. Also, pedagogically relevant was the finding that WM linked only to some discrete areas of performance, but not to the holistic score of L2 writing quality. These results resonate with the findings in Bulté and Housen (2014), who reported that the objective CAF measures, which reflected L2 writing development over time, did not correlate with the subjective ratings given by human raters. This disjunction between the holistic ratings and more objective quantitative measures points to the necessity to reconsider assessment practices. In particular, holistic grading by means of rubrics could be complemented with the discrete and objective assessment of the CAF dimensions. In our view, such refinement of the assessment practices will provide teachers with a clearer and more nuanced picture of L2 writing performance and will be instrumental for the development of more targeted and effective teaching strategies.

It is also important to recognize the limitations of this study. The fact that the participants represented a specific group of L2 English learners in a specific learning context limits the generalizability of our findings. Also, our participants belonged to the upper-intermediate and advanced levels of L2 proficiency. Consequently, caution is required when applying these findings to beginning-level L2 writers. Another consideration concerns the fact that the experiment was carried out in laboratory conditions, which might have affected the participants' writing behaviours. Also, our WM test did not distinguish between different sub-dimensions of WM (phonological memory, task-switching, updating, inhibition), which limits our understanding of the specific contribution of various WM subcomponents to L2 writing production. Additionally, there are certain limitations in L2 writing assessment. In particular, only one measure of fluency (length of T-unit) was employed. We must stress there are dual views on the use of this measure in the assessment of L2 performance. Thus, while some researchers consider this measure to tap into syntactic complexity (Norris & Ortega, 2009), Wolfe-Quintero et al. (1998) consider the length of the T-unit to be one of "the best measures"of writing fluency (p. 29). An important limitation, however, is the fact that this length-based measure does not shed light on the production processes that enable a writer to compose a text more fluently. Future research should strive for the multidimensional analysis of fluency, combining product-based measures (length of T-unit) with the measures of composing rate (syllables per minute) and also process-based measures which draw upon the online observations of L2 writers' composing processes (e.g., length of text between pauses) (Abdel Latif, 2013). Another consideration concerns the assessment of propositional complexity. Albeit supplemented with the comprehensive syntactic criteria, the primary requirement of the guidelines is to determine the boundaries of idea units based on the "semantic integrity" and "meaningfulness" of a discourse chunk. And despite the clear definition of these semantic criteria in the guidelines, the segmentation of discourse into idea units involves a certain amount of subjectivity. Thus, more empirical evidence is required concerning the validity of the idea unit guidelines.

In spite of these limitations, the findings from this study hopefully enhance our knowledge of the mechanism by which WM intervenes in the process of L2 writing. We hope that this study will motivate more research involving the participants of all levels of L2 proficiency, L1 and L2 backgrounds, and various learning contexts. Future work could also explore the role of cognitive abilities in L2 writing from a process-oriented perspective, employing such techniques as eye-tracking, key-stroke logging, and stimulated recall. Also, more research about the trainability of WM is needed, as we need to know more about whether the training of WM may have direct benefits for L2 writing skills.

Funding

This study was supported by the National Research Agency (grants PID2019-110594GB-100 and PID2019-104353GB-I00) and the grant from the Seneca Foundation (grant No. 20832/PI/18).

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

The research reported in this paper is part of the work conducted within two research projects on L2 writing financed by the Spanish *Ministerio de Economía y Competitividad* (research grant FFI 2016-79763-P) and by *Fundación Séneca* (research grant 20832/PI/18).

References

Abdel Latif, M. M. (2013). What do we mean by writing fluency and how can it be validly measured? *Applied Linguistics, 34*(1), 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*(3), 288–318. https://doi.org/10.1037/0096-3445.117.3.288.

Adams, A. M., & Gathercole, S. E. (1995). Phonological working memory and speech production in preschool children. *Journal of Speech and Hearing Research, 38*(2), 403–414. https://doi.org/10.1044/jshr.3802.403.

Adams, A. M., & Guillot, K. (2008). Working memory and writing in bilingual students. *ITL - International Journal of Applied Linguistics, 156*(1), 13–28. https://doi.org/10.2143/ITL.156.0.2034417.

Ahmaddian, M., & Vasylets, O. (2021). The role of cognitive individual differences in writing performance and written corrective feedback processing and use. In R. M. Manchón, & C. Polio (Eds.), *Handbook of second language acquisition and writing*. Routledge.

Alptekin, C., Erçetin, G., & Özemir, O. (2014). Effects of variations in reading span task design on the relationship between working memory capacity and second language reading. *Modern Language Journal, 98*(2), 536–552. https://doi.org/10.1111/modl.12089.

Baaijen, V. M., Galbraith, D., & de Glopper, K. (2014). Effects of writing belifs and planning on writing performance. *Learning and Instuction, 33*, 81–91. https://doi.org/10.1016/j.learninstruc.2014.04.001.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829–839. https://doi.org/10.1038/nrn1201.

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. https://doi.org/10.21832/9781783093595.

Baddeley, A., & Hitch, G. (1974). Working memory. *The Psychology of Learning and Motivation, 8*, 47–89. https://doi.org/10.1016/S0079-7421(08)60452-1.

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

Bergsleithner, J. M. (2010). Working memory capacity and L2 writing performance. *Ciências & Cognição, 15*(2), 2–20. https://bit.ly/30JOsrh.

Berman, R., & Slobin, D. (Eds.). (2013). *Relating events in narrative: A crosslinguistic developmental study*. Lawrence Erlbaum.

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21–47). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.32.

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing, 26*, 42–65. https://doi.org/10.1016/j.jslw.2014.09.005.

Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. University of Chicago Press.

Chai, W. J., Abd Hamid, A. I., & Abdullah, J. M. (2018). Working memory from the psychological and neurosciences perspectives: A review. *Frontiers in Psychology, 9*, 1–16. https://doi.org/10.3389/fpsyg.2018.00401.

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

Conway, A. R., 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*(5), 769–786. https://doi.org/10.3758/BF03196772.

Council of Europe, Council for Cultural Cooperation, Education Committee, & Modern Languages Division. (2001). *Common European framework of reference for languages*. Cambridge University Press.

Cowan, N. (2005). *Working memory capacity*. Psychology Press.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*(4), 422–433. https://doi.org/10.3758/BF03214546.

Dussias, P. E., & Piñar, P. (2010). Effects of reading span and plausibility in the reanalysis of wh-gaps by Chinese-English second language speakers. *Second Language Research, 26*(4), 443–472. https://doi.org/10.1177/0267658310373326.

Field, A. (2009). *Discovering statistics using SPSS*. Sage.

Gathercole, S., & Alloway, T. P. (2008). *Working memory and learning: A practical guide for teachers*. Sage Publications.

Gilabert, 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.

Grañena, G., Jackson, D. O., & Yilmaz, Y. (Eds.). (2016). *Cognitive individual differences in second language processing and acquisition*. John Benjamins Publishing Company. https://doi.org/10.1075/bpa.3.

Grey, S., Cox, J. G., Serafini, E. J., & Sanz, C. (2015). The role of individual differences in the study abroad context: Cognitive capacity and language development during short-term intensive language exposure. *Modern Language Journal, 99*(1), 137–157. https://doi.org/10.1111/modl.12190.

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*(4), 339–387. https://doi.org/10.1006/cogp.2001.0769.

Hayes, A. F. (2018). *Introduction to mediation. A regression-based approach*. Guilford Publications, Inc.

Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication, 29*(3), 369–388. https://doi.org/10.1177/0741088312451260.

Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). *Range [Computer software]*. Retrieved from http://RANGE HOME (lextutor.ca).

Hoskyn, M., & Swanson, H. L. (2003). The relationship between working memory and writing in younger and older adults. *Reading and Writing, 16*(8), 759–784. https://doi.org/10.1023/A:1027320226283.

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

Housen, A., Kuiken, F., & Vedder, I. (2012). *Dimensions of L2 performance and proficiency*. John Benjamins Publishing Company.

Joh, J., & Plakans, L. (2017). Working memory in L2 reading comprehension: The influence of prior knowledge. *System, 70*, 107–120. https://doi.org/10.1016/j.system.2017.07.007.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching, 44*(2), 137–166. https://doi.org/10.1017/S0261444810000509.

Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy, & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 57–71). Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203811122.

Kellogg, R. T., Whiteford, A. P., Turner, C. E., Cahill, M., & Mertens, A. (2013). Working memory in written composition: An evaluation of the 1996 model. *Journal of Writing Research, 5*(2), 159–190. https://doi.org/10.17239/jowr-2013.05.02.1.

Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing, 20*, 148–161. https://doi.org/10.1016/j.jslw.2011.02.001.

Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing, 21*(4), 390–403. https://doi.org/10.1016/j.jslw.2012.09.003.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism, 11*(2), 261–271. https://doi.org/10.1017/S1366728908003416.

Kuiken, F., Vedder, I., & Gilabert, R. (2010). Communicative adequacy and linguistic complexity in L2 writing. In I. Bartning, M. Martin, & I. Vedder (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 81–100). Eurolsa Monograph Series.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics, 27*(4), 590–619. https://doi.org/10.1093/applin/aml029.

Li, S. (2015). The associations between language aptitude and second language grammar acquisition: A meta-analytic review of five decades of research. *Applied Linguistics, 36*(3), 385–408. https://doi.org/10.1093/applin/amu054.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review, 21*(4), 861–883. https://doi.org/10.3758/s13423-013-0565-2.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15*(4), 474–496. https://doi.org/10.1075/ijcl.15.4.02lu.
Lu, Y. (2015). Working memory, cognitive resources and L2 writing performance. In Z. E. Wen, M. B. Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 175–189). Multilingual Matters. https://doi.org/10.21832/9781783095735.
Mackey, A., & Sachs, R. (2012). Older learners in SLA research: A first look at working memory, feedback, and L2 development. *Language Learning, 62*(3), 704–740. https://doi.org/10.1111/j.1467-9922.2011.00649.x.
Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development*. Palgrave Macmillan. https://doi.org/10.1057/9780230511804.
Manchón, R. M., & Vasylets, O. (2019). Language learning through writing: Theoretical perspectives and empirical evidence. In A. Benati, & J. W. Schwieter (Eds.), *The Cambridge handbook of language learning* (pp. 341–363). Cambridge University Press. https://doi.org/10.1017/97811083333603.
Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition, 34*(3), 379–413. https://doi.org/10.1017/S0272263112000125.
Mavrou, I. (2020). Working memory, executive functions, and emotional intelligence in second language writing. *Journal of Second Language Writing, 50*. https://doi.org/10.1016/j.jslw.2020.100758.
McCutchen, D. (2011). From novice to expert: Implications of language skills and writing-relevant knowledge for memory during the development of writing skill. *Journal of Writing Research, 3*(1), 51–68. https://doi.org/10.17239/jowr-2011.03.01.3.
Meara, P. M., & Miralpeix, I. (2006). *Y_Lex: The Swansea advanced vocabulary test*. Lognostics.
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.
Murphy, K., & Creux, O. (2020). Examining the association between media multitasking, and performance on working memory and inhibition tasks. *Computers in Human Behavior, 114*. https://doi.org/10.1016/j.chb.2020.106532.
Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*(4), 555–578. https://doi.org/10.1093/applin/amp044.
O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524490.
Olive, T. (2012). Working memory in writing. In V. W. Berninger (Ed.), *Past, present, and future contributions of cognitive writing research to cognitive psychology* (pp. 485–503). Taylor & Francis. https://doi.org/10.4324/9780203805312.
Ong, J., & Zhang, L. J. (2010). Effects of task complexity on fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing, 19*(4), 218–233. https://doi.org/10.1016/j.jslw.2010.10.003.
Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition, 21*(1), 109–148. https://doi.org/10.1017/s0272263199001047.
Pollard, M. A., & Courage, M. L. (2017). Working memory capacity predicts effective multitasking. *Computers in Human Behavior, 76*, 450–462. https://doi.org/10.1016/j.chb.2017.08.008.
Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics, 25*, 46–73. https://doi.org/10.1002/9781405198431.wbeal0035.
Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua, 120*(8), 2022–2039. https://doi.org/10.1016/j.lingua.2010.02.004.
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*(1), 31–79. https://doi.org/10.1111/j.1467-9922.2010.00590.x.
Schwaighofer, M., Fischer, F., & Bühner, M. (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educational Psychologist, 50*(2), 138–166. https://doi.org/10.1080/00461520.2015.1036274.
Serafini, E. J., & Sanz, C. (2016). Evidence for the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition, 38*(4), 607–646. https://doi.org/10.1017/S0272263115000327.
Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences in instructed language learning* (pp. 69–95). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.2.06ske.
Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research, 1*(3), 185–211. https://doi.org/10.1177/136216889700100302.
Tsai, N., Au, J., & Jaeggi, S. M. (2016). Working memory, language processing, and implications of malleability for second language acquisition. In G. Grañena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 69–88). John Benjamins Publishing Company. https://doi.org/10.1075/bpa.3.
UCLES. (2001). *Oxford quick placement test*. Oxford University Press.
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.
Vanderberg, R., & Swanson, H. L. (2007). Which components of working memory are important in the writing process? *Reading and Writing, 20*(7), 721–752. https://doi.org/10.1007/s11145-006-9046-6.
Vasylets, O. (2017). *Task modality effects: A study of task complexity effects in speech and writing*. Barcelona, Spain: University of Barcelona [Doctoral thesis] http://hdl.handle.net/2445/113218.
Vasylets, O., Gilabert, R., & Manchón, R. M. (2017). The effects of mode and task complexity on second language production. *Language Learning, 67*(2), 394–430. https://doi.org/10.1111/lang.12228.
Vasylets, O., Gilabert, R., & Manchón, R. M. (2019). Differential contribution of oral and written modes to linguistic and propositional complexity in L2 performance. *Instructed Second Language Acquisition, 3*(2), 206–227. https://doi.org/10.1558/isla.38289.
Vasylets, O., Gilabert, R., & Manchón, R. M. (2020). Task modality, communicative adequacy and CAF measures: The moderating role of task complexity. In R. M. Manchón (Ed.), *Writing and language learning: Advancing research agendas* (pp. 183–207). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.56.08vas.
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.
Wen, Z. (2015). Working memory in second language acquisition and processing: The phonological/executive model. In Z. E. Wen, M. B. Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing*. Multilingual Matters. https://doi.org/10.21832/9781783093595.
Wen, Z., Biedroń, A., & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching, 50*(1), 1–31. https://doi.org/10.1017/S0261444816000276.
Williams, J. (2012). Working memory and SLA. In S. M. Gass, & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 427–441). Routledge. https://doi.org/10.4324/9780203808184.
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 Hawaii Press.
Zabihi, R. (2018). The role of cognitive and affective factors in measures of L2 writing. *Written Communication, 35*(1), 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*(2), 335–352. https://doi.org/10.1111/modl.12389.

Olena Vasylets is an Associate Professor at the University of Barcelona, Spain. Her principle research interests are cognitive processes in language learning, the role of mode (oral, written and multimodality) in second language acquisition, and computer-assisted language learning. She has published articles in peer-reviewed journals such as Language Learning and Annual Review of Applied Linguistics.

Javier Marín is a Full-time Professor at the University of Murcia, Spain. His main research interests interest are psycholinguistic processes involved in language learning and learning disabilities. He has published extensively in edited volumes and peer-reviewed journals.