The role of cognitive individual differences in digital versus pen-and-paper writing

Olena Vasylets
University of Barcelona, Spain
https://orcid.org/0000-0003-0241-1279
vasylets@ub.edu

M. Dolores Mellado
University of Murcia, Spain
https://orcid.org/0000-0002-9089-0152
mariadolores.mellado@um.es

Luke Plonsky
Northern Arizona University, Flagstaff, USA
https://orcid.org/0000-0002-5791-1839
lukeplonsky@gmail.com

Abstract
It is unknown whether and to what extent cognitive individual differences may play different roles in paper versus computer-based second language (L2) writing. This exploratory study is a first attempt to explore this issue, focusing on the effects of working memory and language aptitude on the quality of paper versus computer-based L2 writing performance. Forty-two Spanish learners of L2 English performed a problem-solving task either digitally or on paper, took a working memory *n*-back test, and completed LLAMA tests to measure language aptitude. The quality of their L2 written texts was assessed in terms of complexity, accuracy and fluency (CAF) measures. The results indicated that the role of cognitive individual differences may vary depending on the writing environment.

Keywords: writing environment; pen-and-paper writing; digital writing; working memory; language aptitude

1. Introduction

Because of its complexity, writing ability develops at a varied pace and is characterized by high variability in ultimate attainment, even in the first language (L1) (Bereiter & Scardamalia, 1987). Second language (L2) writing can be even more complicated due to additional challenges, such as gaps in L2 knowledge or lack of automatization of L2 spelling rules (Weigle, 2005). Variability in L2 writing can also be attributed to individual differences, both cognitive and affective, that learners bring into the writing task (Kormos, 2012, 2023; Papi et al., 2022). The mixed nature of empirical evidence, however, precludes a nuanced understanding of the role of cognitive resources in L2 writing. The diversity of the research findings could be attributed, inter alia, to learner internal and external factors which can moderate the relationship between cognitive individual differences and L2 writing. Importantly, writing represents a highly embodied activity, in the sense that it is contingent on the interactions between the writer's mind, body and external environment (Mangen & Balsvik, 2016). The writing environment (i.e., handwriting on paper versus typing using the computer), thus, represents a learner-external factor which is central to writing activity.

The effects of environment in writing can be justified from multiple theoretical perspectives, (e.g., Hayes, 2012; Kress, 2003; Mangen & Velay, 2010), and findings from neuroscience (Askvik et al., 2020; Ihara et al., 2021) and writing research (Chan et al., 2017) have shown that the nature of learning and performance can differ depending on whether paper or computer is involved. Surprisingly, however, except for the research on testing (Barkaoui & Knouzi, 2018), the performance environment has been practically ignored in second language acquisition (SLA) research. Recently, however, writing environment has been defined as a task complexity factor (Vasylets & Marín, 2022), which gives further theoretical justification for the hypothesis of a differential involvement of individual differences in paper versus digital writing (Robinson, 2011). However, the empirical evidence to substantiate these claims is still lacking. The neglect of the role of environment in empirical research is exemplified by the fact that recent meta-analyses have not assessed whether and to what extent environment might moderate the relationship between language aptitude and L2 proficiency (Li, 2016) or between working memory and L2 reading (Jeon & Yamashita, 2014; Peng et al., 2018; Shin, 2020). A rare exception is found in In'nami et al.'s (2022) study, which observed stronger correlations between working memory and paper-based reading tasks as compared to computer-based tasks. This finding provides a tentative indication that the relationship between cognitive abilities and L2 processing and outcomes might be moderated by the environment of task performance. To gain a deeper understanding of this issue, this study explores

whether the writing environment might moderate the effects of working memory and language aptitude on L2 writing performance.

2. Literature review

2.1. Pen-and-paper versus computer-based writing

Pen-and-paper and computer-based writing present important differences in the way writers use their own body and interact with the external environment during text production (Clark, 2001; Clark & Chalmers, 1998). The most notable difference between the two writing environments lies in the transcription processes. Thus, in pen-and-paper writing some sort of stylus is employed to hand-craft written signs on paper. On the other hand, in computer writing a keyboard is used to select ready-made written signs which appear on the screen. Reading behaviors also differ as the fixed layout and tangible nature of paper is believed to benefit stable and efficient visual representation of the written text (Hou et al., 2017). As such, pen-and-paper writing represents a rich kinesthetic and haptic experience which is also laborious and slow (Mangen, 2016). On the other hand, computer-writing is faster and less laborious. However, the use of keyboard and screen is also believed to convert digital writing into a detached and mediated experience, which is more phenomenologically monotonous and impersonalized than pen-and-paper writing (Kiefer et al., 2015).

From a purely conceptual standpoint, the role of the environment in writing is acknowledged in multiple theoretical perspectives. Thus, transcription processes form part of all relevant cognitive models of writing (Flower & Hayes, 1980; Kellogg, 1996). Importantly, in the recent update of his writing model, Hayes (2012) incorporated the element of *transcribing technology*, although without providing any testable predictions concerning its role in writing performance and learning. In a similar vein, a major theoretician in semiotics, Kress (2003, p. 3), points out important changes that digital technology produces in writing:

> The combined effects on writing of the dominance of the mode of image and of the medium of screen will produce deep changes in the forms and functions of writing. This in turn will have profound effects on human, cognitive/affective, cultural and bodily engagement with the world, and on forms and shapes of knowledge.

Theoretical justification for the role of environment can also be found in the tenets of embodied cognition (Clark, 2001; Clark & Chalmers, 1998; Wilson & Golonka, 2013). Conceptions of embodiment take many forms (Barsalou, 2008), but the main underlying idea is that cognition represents a combination

of multiple resources, which include mind, body and their relations to the external world. A major proponent of embodiment in SLA is Atkinson (2011), who introduced the socio-cognitive perspective as an alternative approach to explain L2 learning. The core claim of this approach is that "mind, body, and world function integratively in second language acquisition" (Atkinson, 2011, p. 143). Although Atkinson (2011) admits that the sociocognitive view is "new and undeveloped" (p. 162), he also stresses that this standpoint is open to the full range of possibilities and applications, including L2 writing (Nishino & Atkinson, 2015). One of the most recent applications of embodied cognition views to L2 writing is found in Vasylets and Marín (2022) who proposed that writing environment can be conceptualized as a task complexity factor, given that paper-based and computer writing can pose different cognitive demands on L2 learners. Following this line of thinking and drawing on Robinson's (2011) prediction that "individual differences in affective and cognitive abilities . . . will increasingly differentiate learning and performance as tasks increase in complexity" (p. 19), we could thus hypothesize that individual differences may play out differently in paper-based versus computer writing.

In terms of the empirical evidence, numerous studies in neuroscience, experimental psychology, and writing have found differences in learning and performance in the two writing environments. Thus, various studies have found the advantage of pen-and-paper writing over computer writing in improving spelling (Cunningham & Stanovich, 1990) as well as letter and word learning (Ihara et al., 2021; Longcamp et al., 2006). In addition, a recent study by Askvik et al. (2020) showed that, as compared to typing, handwriting was associated with increased activation in the brain areas important for memory and learning. Rich haptic-kinesthetic experience, which is believed to facilitate encoding of new information, is the common explanatory factor of the learning advantage of pen-and-paper writing.

There is also empirical evidence (albeit mixed) to show that writing processes (Chan et al., 2017) and performance quality (Cheung, 2012) may differ in the two writing environments. For example, Chan et al. (2017) reported that the participants felt more comfortable when planning and revising using the computer; at the same time, writers were more careful during linguistic formulation in paper writing, which was attributed to the difficulty to make changes in handwritten texts. Similarly, participants in Zhi and Huang (2021) reported that they perceived their writing processes to be more authentic in computer writing, while the inconvenience of revision in paper writing induced them to modify their natural writing behaviors.

In sum, there is empirical evidence from various fields that shows that learning affordances, writing processes and performance may differ between pen-and-paper and computer writing. This provides the basis for the empirical justification of the hypothesis that writers may employ their cognitive resources differently depending on the writing environment.

2.2. Language aptitude in L2 writing

Foreign language aptitude, which is generally defined as a specific talent for learning a foreign or second language (Carroll, 1981; Skehan, 2002), is recognized as one of the central cognitive abilities in language learning in general (Wen et al., 2017) and in L2 writing in particular (Ahmaddian & Vasylets, 2022). Since its inception, the construct of language aptitude has been recognized as multicomponential (Dörnyei, 2005). Thus, the classical framework by Carroll (1981) identifies four components of language aptitude: (1) phonetic coding ability, which consists in the ability to learn sound-symbol associations; (2) grammatical sensitivity, which refers to the ability to identify grammatical functions of words; (3) rote learning ability, which is an ability to learn sound-meaning associations; and (4) deductive learning ability, which refers to the ability to induce language rules from input. Recent theoretical work has proposed that the role of language aptitude in SLA may be rather intricate and task/instruction-specific (Robinson, 2005; Skehan, 2002). Similarly, Dörnyei (2010) defined aptitude as a complex system which dynamically interacts with the learning environment and can, thus, be affected by learner internal and external factors (see also Grañena, 2013; Kormos, 2013).

In this line of thinking, Kormos (2012) hypothesized the specific effects that aptitude components may have on L2 writing. Thus, phonetic coding ability is expected to contribute to more accurate spelling; higher levels of grammatical sensitivity and deductive ability are expected to benefit linguistic encoding; rote ability can benefit lexical complexity of L2 writing. Also, learners with high rote ability, who can potentially have a richer vocabulary, could be expected to produce more lexically complex written texts. Finally, good deductive skills are predicted to help learners handle the grammatical encoding of the conceptual plan more efficiently.

Although the role of aptitude in L2 writing has a solid theoretical justification, empirical findings are scarce and inconclusive. Thus, while aptitude appeared as a strong predictor of general L2 proficiency in Li`s (2016) meta-analysis ($r = .49$, 95%; $CI = .45$-$.54$), findings for writing were not statistically significant, except for two aptitude components related to number learning and spelling clues. Li (2016) explains these unexpected results by the fact that writing might require a different set of skills from those measured in traditional aptitude tests. However, it must be mentioned that in his analysis Li did not consider the potential mediating role of the writing environment in the aptitude effects. By looking at the individual empirical studies, we can observe that writing environment has never been considered as an important variable, to the extent that some studies do not even explicitly indicate it. For example, the oft-cited study by Kormos and Trebits (2012) found that learners with high grammatical sensitivity produced longer clauses in the task which was more demanding in terms

of linguistic encoding; however, no relationship between quality of written production and aptitude was found in the task which posed high demands on content conceptualization. Another recent study by Yang et al. (2019) showed that L2 writing quality, as assessed by a holistic score, was predicted by vocabulary learning and grammar inferencing abilities, which are believed to tap into aptitude for explicit language learning (Grañena, 2013). Importantly, neither of the above-mentioned studies has specified the environment of writing task performance. Absence of this information limits generalizability of the research findings and precludes a more fine-grained understanding of the effects of aptitude on L2 writing.

2.3. Working memory in L2 writing

Working memory (WM) represents another cognitive trait posited to be important both in L1 (Hayes, 2012; Kellogg, 1996) and L2 writing (Ahmaddian & Vasylets, 2022; Kormos, 2012; Papi et al., 2022). WM represents a limited cognitive system responsible for the maintenance in active attention of the task relevant information and inhibition of irrelevant information (Baddeley, 2003).

 Cognitive models of writing by Kellogg (1996) and Hayes (1996, 2012) posit that WM plays a central role in writing. Importantly, while Hayes (1996) considers that all writing processes rely on WM resources, Kellogg contemplates the involvement of WM only in high-level processes of planning, linguistic encoding and monitoring. Taking a perspective of automaticity theories (Schneider & Shiffrin, 1977), it could be argued, however, that execution processes (typing and handwriting) could also draw on WM resources if they are not sufficiently automatized (see also, Chenoweth & Hayes, 2001). Thus, as pointed out by Kormos (2012), typing and handwriting would also draw on WM resources unless they are fully automatized. Considering that the same writer may have different levels of automatization of typing versus handwriting skills, we could also hypothesize that WM resources could be differentially involved, depending on the environment of writing performance. This assumption, however, still needs empirical verification. The available evidence in L1 writing has largely shown positive correlations between WM and L1 writing quality of writers across different ages (Hoskyn & Swanson, 2003; Vanderberg & Swanson, 2007). This supporting evidence, however, comes largely from writing on paper, so it is not clear if the writing environment moderates the relationship in WM and L1 writing.

 Because of potential gaps in L2 linguistic knowledge and/or lack of automatization of orthographic rules, L2 writers might face even greater challenges (Weigle, 2005). For example, less proficient L2 writers might require WM resources for encoding procedures as well as for text monitoring/reviewing. Importantly, transcribing processes (i.e., handwriting and typing) might demand

attention if the spelling rules have not yet been automatized or if the orthography of a writer's L1 is substantially different from that of the L2 (Kormos, 2012). Taking into account that L2 transcribing can be resource-demanding, we could hypothesize that the involvement of WM in L2 writing can vary depending on the environment of performance. The available empirical evidence, however, does not allow for verification of this hypothesis. Thus, the meta-analysis by Linck et al. (2014) reported a positive correlation between WM and L2 writing outcomes, with the estimated population effect size (*p*) of .255. This meta-analysis, however, did not consider writing environment as a potential moderating factor. Examination of the individual studies also reveals that, similar to the research on language aptitude, studies on WM in L2 writing have never considered writing environment as a relevant factor, with some investigations even failing to report it (see Table 1).

Table 1 Studies exploring the relationship between working memory and L2 writing quality

Study	L2	Participants, mean age	Writing environment	Relationship between WM and writing quality
Adams & Guillot (2008)	French/English	12	Paper	Partially positive
Mavrou (2020)	Spanish	20	Paper	Partially positive
Cho (2018)	English	20	Computer	Null
Michel et al. (2019)	English	12	Computer	Null
Vasylets & Marín (2021)	English	19	Computer	Partially positive
Kormos & Sáfár (2008)	English	15-16	Not indicated	Partially positive
Lu (2015)	English	20	Not indicated	Null
Zabihi (2018)	English	21	Not indicated	Mixed

Another conclusion which can be drawn is the mixed nature of previous results. This attests to the complex nature of the relationship between WM and L2 writing performance, giving evidence to Williams's (2015) contention that "the relationship between WM capacity and L2 processing and learning is far more complex and nuanced than originally envisaged" (p. 301; see also Baddeley, 2015). The potential effects of WM can be complexified by the moderating influence of learner internal and external factors. A recent study by Vasylets and Marín (2021), for example, showed that the effects of WM on L2 writing was moderated by the level of L2 proficiency, such that at low levels of proficiency WM had a positive association with writing accuracy, while at high levels of proficiency there was a positive link with lexical sophistication. By the same token, we could suggest that the writing environment could influence the way learners draw on WM resources during L2 writing performance (see also In'nami et al., 2022 for the findings in reading). Exploration of this issue would help gain a more nuanced understanding of the intricacy in the links between WM and L2 writing.

Taking into account the identified research gaps, the following research questions guided the present study:

1. Does language aptitude play the same role in L2 writing performance depending on the environment (paper vs. digital) in which a task is performed? (RQ1)
2. Does working memory play the same role in L2 writing performance depending on the environment (paper vs. digital) in which a task is performed? (RQ2)

Given the differences in the nature of the two writing environments and findings from previous research (e.g., In'nami et al., 2022), we hypothesized that working memory and language aptitude would be differentially involved in L2 writing performance in paper-based versus computer-based writing. We wish to emphasize, however, that this hypothesis is non-directional, that is, we make no predictions regarding a stronger relationship in either writing environment.

3. Method

3.1. Participants

A total of 42 native Spanish EFL learners (age: $M = 21.52$; $SD = 1.27$) participated in the present study. The participants were fourth-year applied linguistics undergraduate students at a Spanish university. For the purposes of the study, the participants were randomly divided into the digital group (DG) (4 males, 20 females) and the pen-and-paper group (P&P) (4 males, 14 females). In order to ensure comparable proficiency between the two groups, participants completed the *Oxford Placement Test*. For logistical reasons, the P&P group took the classical version of the *Oxford Placement Test* (Allen, 1992) with a max score of 100, and the DG took the quick *Oxford Placement Test* (UCLES, 2001) with a max score of 60. In order to check whether the two groups were comparable in terms of proficiency, we multiplied the P&P group's scores by .6. The scores of the two groups were then compared using descriptive statistics as well as an independent samples *t*-test. As shown in Table 2, the mean scores of the two groups were very similar, the difference between them was not statistically significant and produced a very small effect size ($d = 0.16$).

Table 2 Comparison of groups on OPT

Group	M (SD)	Difference between groups	
		t (p value)	Cohen's d (95% CI)
Paper and Pencil	47.90 (4.51)	0.53 (.60)	0.16 (-0.45, 0.78)
Digital	46.96 (6.48)		

3.2. Instruments

3.2.1. Measure of working memory capacity

Working memory capacity was assessed by means of an *n*-back test (Kane et al., 2007). In this test the participants were required to press the M key on the computer keyboard if the stimulus (letter) shown on the screen coincided with the stimulus shown three trials ago (3-back task); if the stimulus did not coincide, the participants had to press the N key. After pressing the key, the participants were presented with the feedback on their performance. The total stimulus set consisted of 15 letters which were presented for 500 milliseconds. Every new stimulus was presented every 3000 milliseconds and the participants had 3 seconds to respond. There were three blocks, each of 25 trials. The test was administered online by means of an experiment created in https://www.psytoolkit.org/. The participants took between 5 to 10 minutes to complete the test.

3.2.2. Measures of language aptitude

To assess language aptitude, we employed the LLAMA tests (Meara, 2005). The LLAMA suite consists of several tests tapping into different sub-dimensions of aptitude: (1) LLAMA_B asks the participants to memorize the associations between shapes and sound combinations; this test is believed to measure learners' ability to learn new words; (2) LLAMA_D is phonetic recognition test which measures how effectively the participant can recognize short segments of oral language to which they have been exposed previously; (3) LLAMA_E measures the ability to learn new sound-symbol associations; the test consists of a set of 22 recorded syllables which the participants have to match to a transliteration of the syllable sounds in an unfamiliar language; (4) LLAMA_F measures the ability to infer the rules of an unknown language (i.e., explicit inductive learning ability); based on a set of pictures and sentences describing these pictures, test-takers have to work out the grammatical rules that operate in the language. The LLAMA tests have been shown to have acceptable internal consistency and stability (Grañena, 2013) and they have been widely used in previous empirical studies (Artieda & Muñoz, 2016; Yang et al., 2019).

3.3. Writing task

The participants were invited to produce a written text in response to the complex version of the "Fire-Chief" task (Gilabert, 2005). This task consists of a problem-solving picture-based writing activity in which participants are presented

with an image of a burning building from which numerous people need to be rescued. The task requires the participants to explain and justify the actions they would take in order to save as many people as possible from the burning building.

3.4. Procedure

There were two 50-minute sessions of data collection for both P&P and DG. During the first session, the participants in the P&P group completed the writing task in the computer lab. Each participant was provided with a task prompt and instructions and with a blank writing sheet on which to write their text. The participants were asked to read the instructions carefully and to familiarize with the picture in order to get an overall idea of the situation in the task before starting to write their compositions. The participants were given 50 minutes to perform the task, but there was no specific word limit. The learners finished the writing task within a range of 12-47 minutes ($M = 26.73$; $SD = 9.91$). During the second session, the P&P group completed the LLAMA tests and the working memory test on the computers at the university lab.

The participants in the DG performed all tasks at home using their personal computers. During the first session, the digital group performed the writing task. The participants received the prompt and instructions by email and they were required to email the completed task to the researcher within the time limit of 50 minutes. The participants were asked not to use dictionaries and any other external sources during task completion. During the second session, the DG received by email the instructions for the LLAMA tests and the link to the working memory test, and they performed the tests on their personal computers.

3.5. Analysis of L2 written production

CAF measures were employed as quantitative indicators of L2 writing performance. To assess accuracy, we calculated the ratio of errors per 100 words (all errors/words x 100). We took into account errors in grammar and vocabulary; spelling and punctuation errors were not counted. Total time (in seconds) and words per minute (total words/total time) were employed as the measures of fluency (Wolfe-Quintero et al., 1998). For lexical complexity, we employed Synlex software (Lu, 2010) to obtain automated measures of lexical density, sophistication and diversity (UBER index). We also employed Synlex to obtain automated measures of syntactic complexity, including mean length of T-unit as a general measure of complexity, coordinate phrases per clause as a measure of coordination, dependent clauses per clause to assess subordination; for nominal complexity, mean length of clause and the ratio of complex nominals per clause were calculated.

3.6. Statistical analyses

Prior to running the main analyses for the present study, descriptive statistics (means and standard deviations) were calculated for all independent and dependent variables. In order to address the two research questions guiding this study, a series of correlations were carried out at the group level between (a) the different measures of working memory (n-back) and aptitude and (b) each of the measures of writing quality for each group. These correlations themselves were not the focus of the study, however. Rather, these correlations were then compared to understand whether and to what extent they might differ across the two writing conditions. Toward that end, a statistical test was conducted to assess whether the difference between the observed correlations for each group was statistically significant. The JASP (Jeffrey's Amazing Statistics Program) statistical software package (JASP Team, 2021) was used for all correlational analyses, and the online tool based on the cocor package in R (http://comparingcorrelations.org/) was used to compare the observed correlation coefficients (Diedenhofen & Musch, 2015). This procedure is rarely employed in applied linguistics but it possesses, we feel, substantial potential to help the field better understand certain types of relationships. Finally, we would like to express a note of caution in interpreting these correlations and the differences between them. The present study is based on a relatively small sample which may, along with error in our measurements, introduce a degree of noise that may also obscure our ability to detect the relationships and differences of interest.

4. Results

Before addressing the research questions directly, we present in Tables 3 and 4 the descriptive statistics for all of the dependent and independent measures, respectively, at the group level. Table 3 presents the descriptives for the 11 dependent measures across four variables: accuracy, fluency, lexical diversity, and complexity. Overall, the two groups are fairly similar. However, there are some differences that are perhaps worth noting. For example, the fluency of the P&P group was higher as they wrote about five more words per minute on average; also, P&P group produced almost twice as many errors as the DG. The DG, by contrast, showed signs of greater complexity in that their writing included a much greater number of dependent clauses per clause than the P&P group; on the other hand, coordination tended to be higher in the P&P group.

Table 3 Descriptive statistics for dependent variables

Variable	Pen and paper (*N* = 18)	Digital (*N* = 24)
	M (*SD*)	*M* (*SD*)
Accuracy		
Ratio of errors x 100	10.58 (3.77)	5.82 (2.75)
Fluency		
Time (in seconds)	1633.33 (465.01)	1582.50 (685.43)
Words per minute	18.40 (3.35)	13.07 (5.72)
Lexical complexity		
Lexical density	0.45 (0.04)	0.44 (0.03)
Lexical sophistication	0.17 (0.04)	0.17 (0.04)
UBER	15.94 (1.09)	15.79 (1.95)
Syntactic complexity		
Mean length of T-unit	0.05 (0.01)	0.04 (0.01)
Mean length of clause	0.09 (0.02)	0.09 (0.02)
Coordinate phrases per clause	0.44 (0.10)	0.25 (0.90)
Dependent clauses per clause	0.24 (0.10)	0.50 (0.09)
Complex nominals per clause	1.16 (0.33)	1.01 (0.17)

There were also some marked differences between the groups on the measures of working memory and aptitude (see Table 4). Although the P&P group's scores were much higher on the LLAMA D and LLAMA F, the DG greatly outperformed their counterparts on the LLAMA E and somewhat outperformed them on the n-back task.

Table 4 Descriptive statistics for independent variables

Variable	Pen and paper (*N* = 18)	Digital (*N* = 24)
	M (*SD*)	*M* (*SD*)
Working memory		
n-back	7.50 (4.45)	9.21 (5.93)
Aptitude		
LLAMA B	79.83 (7.51)	70.65 (23.31)
LLAMA D	57.78 (25.33)	33.48 (15.77)
LLAMA E	34.17 (20.88)	81.74 (43.26)
LLAMA F	81.11 (22.40)	42.26 (28.75)

The correlations between our aptitude and CAF measures for the P&P and DG groups are presented in Tables 5 and 6, respectively. We will now explore and compare those correlations in order to address RQ1.

Table 5 Correlations between language aptitude scores and CAF measures in pen-and-paper writing condition

Variable	LLAMA B	LLAMA D	LLAMA E	LLAMA F
Accuracy				
Ratio of errors x 100	-.09	-.16	-.39	-.56*
Fluency				
Time (in seconds)	-.03	-.37	-.36	.23
Words per minute	-.14	-.01	.24	.26
Lexical complexity				
Lexical density	.11	-.18	-.002	-.24
Lexical sophistication	.10	.05	.15	.29
UBER	-.07	.59	.16	-.04
Syntactic complexity				
Mean length of T-unit	.17	.18	-.01	.17
Mean length of clause	.20	.22	-.06	-.04
Coordinate phrases per clause	-.23	-.11	-.05	-.18
Dependent clauses per clause	.03	-.01	.08	.41*
Complex nominals per clause	.03	-.01	.08	.41

Table 6 Correlations between language aptitude scores and CAF measures in digital writing condition

Variable	LLAMA B	LLAMA D	LLAMA E	LLAMA F
Accuracy				
Ratio of errors x 100	-.17	.03	-.40	.12*
Fluency				
Time (in seconds)	.26	-.12	-.18	.08
Words per minute	-.19	-.12	.43	-.03
Lexical complexity				
Lexical density	.16	-.20	.44	-.17
Lexical sophistication	-.48	.21	-.45	.16
UBER	.29	.11	.22	-.01
Syntactic complexity				
Mean length of T-unit	.37	-.14	-.07	.54
Mean length of clause	.19	-.12	-.16	.27
Coordinate phrases per clause	-.09	.23	-.04	-.19
Dependent clauses per clause	-.45	-.21	.05	-.28*
Complex nominal per clause	-.17	-.04	.18	.01

Note. N = 18 for P&P, 24 for DG; * = differences between correlations significant at .05 level

The magnitude of the relationships we observed for the number of errors (measure of accuracy) was small to moderate across measures and across the two groups (see Plonsky & Oswald, 2014, for a set of benchmarks for interpreting correlations in L2 research). As shown in Table 5, the correlations for the LLAMA B and D were almost negligible. The other measures were more moderately correlated with the number of errors. Of note, whereas the P&P group's LLAMA F score was substantially and negatively correlated with the number of

errors (r = -.56), the correlation for the DG was small and positive (r = .12). The difference in this pair of correlations represents one of the few in our results that was found to be statistically significant.

We employed two measures of fluency: time (in seconds) and words per minute (WPM). In contrast to the results obtained for accuracy, the findings for fluency show greater consistency across the two conditions. In other words, very few pairs of correlations exhibited differences in magnitude or direction, none of which were statistically significant.

Three measures of lexical complexity (i.e., lexical density, sophistication and variety) were employed in the present study. The correlations between these variables and the various measures of aptitude ranged from moderate and negative (e.g., r = -.48 for lexical diversity x LLAMA B for the DG) to similarly moderate and positive (e.g., r = .44 for lexical density x LLAMA E for the DG). Several correlations exhibited differences between the two groups in terms of size and/or magnitude (e.g., .59 [P&P] vs. .11 [DG] for UBER x LLAMA D). None of these differences were found to be statistically significant. However, they are no less noteworthy given the span of correlations seen here between the groups.

Five different measures of syntactic complexity were assessed in the present study, each of which was correlated with the four aptitude measures. The majority of the two groups' correlations were fairly similar. However, as with the other measures of L2 writing we have seen thus far, there were a few noteworthy differences in correlation size and/or direction. The correlation between the LLAMA F and the number of complex nominals per clause (CNom/C) was much stronger for the P&P group than the DG (r = .41 vs. .01). And in the case of the correlation between the LLAMA F and the number of dependent clauses per clause (DepC/C), the correlation was not only weaker for DG but also negative (r = -.28), compared to a stronger and positive correlation for the P&P group (r = .41), a difference that was found to be statistically significant.

5. Discussion

The aim of this study was to explore whether and to what extent language aptitude and working memory were similarly or differently involved in paper-based versus digital writing. Based on previous theorizing (e.g., Mangen & Velay, 2010) as well empirical findings (In'nami et al., 2022), we tentatively hypothesized that the role of cognitive individual differences in L2 written performance would vary depending on the environment of production. To test this hypothesis, we conducted a study in which a group of Spanish EFL university learners of advanced L2 proficiency took a working memory test (n-back), a language aptitude test (LLAMA tests), and performed a problem-solving task either digitally or on paper.

In the first place, the comparison of the CAF measures showed some differences in the quality of paper and digital texts. Thus, we found that accuracy was higher in the DG, as the P&P group produced twice as many errors. This finding for accuracy can be explained by the fact the learners in the DG could benefit from the spell-checkers. Notably, learners spent roughly the same time on the text production in both conditions. However, speed fluency appeared to be higher in the P&P group who wrote about five more words per minute on average. At the first sight, this finding may seem counterintuitive, as computer writing is inherently faster. To explain this finding, we can tentatively suggest that learners in the DG revised and edited their text to a higher extent, which eventually resulted in slower speed fluency as measured by the number of words per minute. This finding resonates with some recent studies (e.g., Chan et al., 2017) which reported that the facility of revision in digital writing induced more intensive revision processes as compared to paper writing. Although lexical complexity was largely similar in the two modalities, there were some differences in terms of syntactic complexity, with higher indices of coordination observed in paper writing, but higher subordination in digital writing. In sum, these results align with previous research which reported that the nature of writing processes as well as writing quality may vary depending on writing environment (Cheung, 2012; Zhi & Huang, 2021).

As for the main research questions, the experimental results partially confirmed our initial hypothesis of a differential involvement of cognitive individual differences in paper versus digital writing. Thus, concerning language aptitude, the two groups were overall very similar in terms of the correlations between CAF measures and LLAMA B, D and E. In fact, the correlations for LLAMA D (phonetic recognition) and LLAMA B (vocabulary learning) were almost negligible, while the correlations for LLAMA E (sound-symbol learning) ranged from small to moderate, but without reaching statistical significance. Notable differences, however, were observed between the two groups in the size and nature of the correlations between CAF measures and LLAMA F, which measures grammar inferencing ability. For example, whereas for the P&P group the correlation between LLAMA F score and the number of errors was substantial and negative (r = -.56), the correlation for the DG was small and positive. Differences were also observed in the area of syntactic complexity. For example, the correlation between the LLAMA F and nominal complexity (the number of complex nominals per clause) was much stronger for the P&P group than the DG. Also, for the DG, the correlation between LLAMA F and subordination (the number of dependent clauses per clause) was weak and negative (r = -.28), whereas there was a stronger and positive correlation for the P&P group (r = .41), a difference that was found to be statistically significant. Our findings for LLAMA F and P&P modality

resonate with the results of previous studies (e.g., Kormos & Trebits, 2012; Yang et al., 2019) which also reported a positive relationship between grammar inferencing ability and quality of L2 writing performance. The notable finding in this study, however, is that the role of grammar inferencing ability may vary depending on the environment (paper versus digital) in which a task is performed.

A similar tendency was also observed for working memory. Thus, one notable finding was that the correlation between working memory and the number of errors was positive and moderate for P&P (r = .26) (for similar findings, see Zabihi, 2018) but negative and moderate for the DG (r = -.31) (see Vasylets & Marín, 2021). Differences in the nature of the correlations in the two writing environments were also observed between working memory scores and some measures of fluency (words per minute) and lexical complexity (density, diversity); notably, the direction of the correlations differed for all measures of syntactic complexity, except for the measure of general syntactic complexity (mean length of T-unit). However, in both writing conditions, the size and magnitude of the correlations were similar, which we consider noteworthy given the different direction of these correlations seen between the groups. Given that some correlations for working memory were negative, our findings partially contradict the results of Linck et al.'s (2014) meta-analysis which reported an overall positive correlation between working memory and L2 writing outcomes. This meta-analysis, however, does not consider writing environment as a potentially moderating factor, which can explain the discrepancies between our and Linck et al.'s (2014) results. The complex pattern of findings for working memory obtained in this study, aligns, however, with the ideas of Williams (2015) and Baddeley (2015), who emphasized nuanced involvement of working memory in SLA performance/production and called for more research striving for a fine-grained understanding of the role of this cognitive resource in SLA.

6. Conclusion

In sum, the findings of this study provide an indication (albeit tentative) that the role of cognitive individual differences in L2 writing may vary depending on the environment (i.e., paper *vs.* digital) in which a task is performed. Inherent differences in the haptic-kinesthetic experiences (richer experiences on paper versus less embodied and detached in digital writing), visual text presentation (stable and tangible on paper versus shifting and dynamic on the screen), as well as the way writing processes are implemented (easy revision/editing on the computer versus complicated revision/editing on paper) can account, inter alia, for the fact that cognitive resources can be differentially involved in paper versus digital writing. More controlled experiments are needed to clarify the mechanisms which account for the variability in the effects of cognitive resources in different environments of L2 writing.

Despite these potential contributions this study makes to our understanding of the role of modality in explaining relationships between individual differences and L2 writing performance, we feel the need to highlight a small number of limitations. In the first place, it should be taken into account that the participants in the pen-and-paper and digital groups obtained different scores on working memory and aptitude tests; additionally, the two groups performed the tasks under different conditions (at home vs. computer lab), which could have had a potential influence on the test/task results. Another important consideration is that we do not have precise estimates of reliability for our independent or dependent measures. Furthermore, any lack of reliability, whether stemming from internal consistency or other sources of non-construct relevant variance, can attenuate our ability to estimate our relationships of interest (see McKay & Plonsky, 2021). Such error may also have contributed to instability in our estimates and a lack of clear differences across writing environments. Future research in this area might consider addressing (i.e., estimating and accounting for) these and other psychometric properties of the measures being employed. Another general concern in the present study relates to our measures of aptitude. Although previous studies have sought to validate the LLAMA battery, further efforts in this area are needed. In particular, the data we collected for the present study did not support the aggregation of the LLAMA subsets into a single aptitude score, thus calling into question the construct validity of the test as a whole (but not necessarily the subtests).

Finally, to our knowledge, this is the first study in applied linguistics to have employed a test for comparing the strength of correlation coefficients (Diedenhofen & Musch, 2015). This procedure may seem unfamiliar but it is not unlike the practice of comparing standardized beta coefficients to understand the relative contributions of different predictors in a multiple regression model (see Mizumoto, in press). We encourage others to consider this technique in cases such as the present when a particular correlation is hypothesized to differ (i.e., be moderated by one or more variables). In spite of the exploratory nature of the study and the tentative nature of our results, we consider that the findings of this study provide the empirical evidence which justifies further comparative research into the role of individual differences in L2 writing.

References

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. (2022). 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* (pp. 139-151). Routledge.

Allen, D. (1992). *Oxford Placement Test 1: Marking kit with user's guide and diagnostic key*. Oxford University Press.

Artieda, G., & Muñoz, C. (2016). The LLAMA tests and the underlying structure of language aptitude at two levels of foreign language proficiency. *Learning and Individual Differences, 50*, 42-48. https://doi.org/10.1016/j.lindif.2016.06.023

Askvik, E., Van der Weel, F. R., & Van der Meer, A. L. (2020). The importance of cursive handwriting over typewriting for learning in the classroom: A high-density EEG study of 12-year-old children and young adults. *Frontiers in Psychology, 11*, 1810. https://doi.org/10.3389/fpsyg.2020.01810

Atkinson, D. (2011). *Alternative approaches to second language acquisition.* Routledge.

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.

Barkaoui, K., & Knouzi, I. (2018). The effects of writing mode and computer ability on L2 test-takers' essay characteristics and scores. *Assessing Writing, 36*, 19-31. https://doi.org/10.1016/j.asw.2018.02.005

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*(1), 617-645. https://doi.org/10.1146/annurev.psych.59.103006.093639

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

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83-118). Newbury House.

Chan, S., Chan, C., Derbie, A. Y., Hui, I., Tan, D., Pang, M., Lau, S., & Fong, K. (2017). Chinese calligraphy writing for augmenting attentional control and working memory of older adults at risk of mild cognitive impairment: A randomized controlled trial. *Journal of Alzheimer's Disease: JAD, 58*(3), 735-746. https://doi.org/10.3233/JAD-170024

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication*, *18*(1), 80-98. https://doi.org/10.1177/0741088301018001004

Cheung, Y. L. (2012). Critical review of recent studies investigating effects of word processing-assisted writing and pen-and-paper writing on the quality of writing and higher-level revisions. *Procedia-Social and Behavioral Sciences*, *46*, 1047-1050. https://doi.org/10.1016/j.sbspro.2012.05.246

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

Clark, A. (2001). *Mindware: An introduction to the philosophy of cognitive science.* Oxford University Press.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7-19. http://www.jstor.org/stable/3328150

Cunningham, A. E., & Stanovich, K. E. (1990). Early spelling acquisition: Writing beats the computer. *Journal of Educational Psychology*, *82*(1), 159-162. https://doi.org/10.1037/0022-0663.82.1.159

Diedenhofen, B., & Musch, J. (2015). cocor: A comprehensive solution for the statistical comparison of correlations. *PLoS ONE, 10*(4). e0121945. https://doi.org/10.1371/journal.pone.0121945

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition.* Lawrence Erlbaum Associates.

Dörnyei, Z. (2010). The relationship between language aptitude and language learning motivation: Individual differences from a dynamic systems perspective. In E. Macaro (Ed.), *Continuum companion to second language acquisition* (pp. 247-267). Continuum.

Flower, L., & Hayes, J. (1980). The dynamics of composing: Making plans and juggling constraints. In L. Gregg & E. Steinberg (Eds.), *Cognitive processes in writing* (pp. 31-50). Lawrence Erlbaum.

Gilabert, 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.

Grañena, G. (2013). Cognitive aptitudes for second language learning and the LLAMA Language Aptitude Test. In G. Grañena & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 105-130). Benjamins.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Randall (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1-27). Erlbaum.

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

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

Hou, J., Rashid, J., & Lee, K. M. (2017). Cognitive map or medium materiality? Reading on paper and screen. *Computers in Human Behavior, 67*, 84-94. https://doi.org/10.1016/j.chb.2016.10.014

Ihara, A., Nakajima, K., Ishimaru, K., Osugi, K., & Naruse, Y. (2021). Advantage of handwriting over typing on learning words: Evidence from an N400 event-related potential index. *Frontiers in Human Neuroscience, 15*, 679191. https://doi.org/10.3389/fnhum.2021.679191

In'nami, Y., Hijikata, Y., & Koizumi, R. (2022). Working memory capacity and L2 reading: A meta-analysis. *Studies in Second Language Acquisition, 44*(2), 381-406. https://doi.org/10.1017/S0272263121000267

JASP Team (2021). JASP (Version 0.15) [Computer software].

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning, 64*(1), 160-212. https://doi.org/10.1111/lang.12034

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 C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 57-71). Lawrence Erlbaum.

Kiefer, M., Schuler, S., Mayer, C., Trumpp, N. M., Hille, K., & Sachse, S. (2015). Handwriting or typewriting? The influence of pen- or keyboard-based writing training on reading and writing performance in preschool children. *Advances in Cognitive Psychology, 11*(4), 136-146. https://doi.org/10.5709/acp-0178-7.

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. (2013). New conceptualizations of language aptitude in second language attainment. In G. Grañena & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 131-152). John Benjamins.

Kormos, J. (2023). The role of cognitive factors in second language writing and writing to learn a second language. *Studies in Second Language Acquisition*.

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

Kormos, J., & Trebits, A. (2012). The role of task complexity, modality, and aptitude in narrative task performance. *Language Learning, 62*(2), 439-472. https://doi.org/10.1111/j.1467-9922.2012.00695.x

Kress, G. (2003). *Literacy in the new media age*. Routledge.

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition*, *38*(4), 801-842. https://doi.org/10.1017/S027226311500042X

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

Longcamp, M., Boucard, C., Gilhodes, J. C., & Velay, J. L. (2006). Remembering the orientation of newly learned characters depends on the associated writing knowledge: A comparison between handwriting and typing. *Human Movement Science*, *25*(4-5), 646-656. https://doi.org/10.1016/j.humov.2006.07.007

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 A. Wen, Z., Mota, M., & Mcneill (Eds.), *Working memory in second language acquisition and processing* (pp. 175-188). Multilingual Matters.

Mangen, A. (2016). What hands may tell us about reading and writing. *Educational Theory*, *66*(4), 457-477. https://doi.org/10.1111/edth.12183.

Mangen, A., & Balsvik, L. (2016). Pen or keyboard in beginning writing instruction? Some perspectives from embodied cognition. *Trends in Neuroscience and Education*, *5*(3), 99-106. https://doi.org/10.1016/j.tine.2016.06.003

Mangen, A., & Velay, J. L. (2010). Digitizing literacy: Reflections on the haptics of writing. In M. H. Zadeh (Ed.), *Advances in haptics* (pp. 386-401). IntechOpen. https://doi.org/10.5772/8710

Mavrou, I. (2020). Working memory, executive functions, and emotional intelligence in second language writing. *Journal of Second Language Writing*, *50*, 100758. https://doi. org/10.1016/j.jslw.2020.100758.

McKay, T., & Plonsky, L. (2021). Reliability analyses: Estimating error in L2 research. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 468-482). Routledge.

Meara, P. M. (2005). *LLAMA language aptitude tests*. 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

Mizumoto, A. (in press). Calculating the relative importance of multiple regression predictor variables using dominance analysis and random forests. *Language Learning*. https://doi.org/10.1111/lang.12518

Nishino, T., & Atkinson, D. (2015). Second language writing as sociocognitive alignment. *Journal of Second Language Writing, 27*, 37-54. https://doi.org/10.1016/j.jslw.2014.11.002

Papi, M., Vasylets, O., & Ahmaddian, M. (2022). Individual difference factors for second language writing. In S. Li, P. Hiver, & M. Papi (Eds.), *Handbook of second language acquisition and individual differences* (pp. 381-395). Routledge.

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., Dardick, W., & Tao, S. (2018). A meta-analysis on the relation between reading and working memory. *Psychological Bulletin, 144*(1), 48-76. https://doi.org/10.1037/bul0000124

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878-912. https://doi.org/10.1111/lang.12079

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics, 25*, 46-73. https://doi.org/10.1017/S0267190505000036

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* (pp. 3-37). John Benjamins.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: II. Detection, search, and attention. *Psychological Review, 84*(1), 1-66. https://doi.org/10.1037/0033-295X.84.1.1

Shin, J. (2020). A meta-analysis of the relationship between working memory and second language reading comprehension: Does task type matter? *Applied Psycholinguistics, 41*(4), 873-900. https://doi.org/10.1017/S0142716420000272

Skehan, P. (2002). Theorizing and updating aptitude. In P. Robinson (Ed.), *Individual differences in instructed language learning* (pp. 69-95). John Benjamins. https://doi.org/10.1075/lllt.2.06ske

UCLES. (2001). *Oxford quick placement test*. Oxford University Press.

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., & Marín, J. (2021). The effects of working memory and L2 proficiency on L2 writing. *Journal of Second Language Writing, 52*. 100786. https://doi.org/10.1016/j.jslw.2020.100786

Vasylets, O., & Marín, J. (2022). Pen-and-paper versus computer-mediated writing modality as a new dimension of task complexity. *Languages, 7*(3), 195. https://doi.org/10.3390/languages7030195

Weigle, S. (2005). Second language writing expertise. In K. Johnson (Ed.), E*xpertise in second language learning and teaching* (pp. 128-149). Palgrave Macmillan. https://doi.org/10.1057/9780230523470_7

Wen, Z. E., 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. (2015). Working memory in SLA research: Challenges and prospects. In Z. E. Wen, M. B. Mota, & A. McNeill (Eds.), W*orking memory in second language acquisition and processing* (pp. 301-308). Multilingual Matters. https://doi.org/10.21832/9781783095735

Wilson, A. D., & Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology, 4*(58). https://doi.org/10.3389/fpsyg.2013.00058

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.

Yang, Y., Sun, Y., Chang, P., & Li, Y. (2019). Exploring the relationship between language aptitude, vocabulary size, and EFL graduate students' L2 writing performance. *TESOL Quarterly*, *53*(3), 845-856. https://doi.org/10.1002/tesq.510

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

Zhi, M., & Huang, B. H. (2021). Investigating the authenticity of computer-and paper-based ESL writing tests. *Assessing Writing*, *50*(1), 100548. https://doi.org/10.1016/j.asw.2021.100548