

## STATE OF THE SCHOLARSHIP

# Working memory and second language writing: A systematic review

Shaofeng Li 

Florida State University, School of Teacher Education, Tallahassee, FL, United States

Email: [sli9@fsu.edu](mailto:sli9@fsu.edu)

(Received 14 September 2022; Revised 18 January 2023; Accepted 10 March 2023)

### Abstract

This article reports on a comprehensive synthesis of the literature on the role of working memory in second language (L2) writing. It starts with an overview and clarification of the construct and measurement of working memory, followed by an elaboration of major theoretical models informing the synthesized research. The article then presents a synthesis of the methods and results of the 16 studies that have been conducted on the associations between working memory and L2 writing. The methodological synthesis encompasses research design, methods of working memory, measurement of writing performance, methods of data elicitation for writing processes, and data analysis and reporting. The results of the synthesized studies demonstrate that (1) working memory is largely unrelated to overall writing proficiency; (2) it is predictive of specific aspects of L2 composition such as complexity, accuracy, and fluency; (3) the role of working memory varies as a function of genre, proficiency, target structure, instruction type, and task demands; and (4) verbal working memory, phonological short-term memory, visual-spatial working memory, and executive functions (inhibiting, shifting, and updating) have differential associations with the process and product aspects of L2 writing. The methods and results are discussed by identifying trends, accounting for disparities, clarifying confusion, recommending solutions, and proposing new directions.

### Introduction

Writing models posit a pivotal role for working memory in the process and product aspects of written composition (Hayes, 1996; Kellogg et al., 2013). The rationale for the importance of working memory in writing is that writing is an effortful process that requires cognitive resources for conscious information processing and that such resources are afforded by working memory, a cognitive space for simultaneous information storage and manipulation. Writing involves the incremental, dynamic, and recursive interaction between information generation, linguistic (phonological, morphosyntactic, and orthographic) encoding, transcription, and editing, which pose a heavy processing demand for writers' working memory resources. The importance of

working memory is supposedly more evident in second language (L2) writing than first language (L1) writing due to the extra cognitive burden caused by L2 writers' incomplete and unautomatized linguistic system and their lack of genre and discoursal knowledge about writing in the L2. Thus, working memory is critical to an accurate understanding of the mechanism and underlying process of L2 writing, and research on working memory has valuable implications for the theory, research, and pedagogy of L2 writing.

This article seeks to synthesize existing research on the associations between working memory and L2 writing and to inspire and inform future research. The article has two major sections: background and research synthesis. The background section provides an overview and clarification on the construct and measurement of working memory, discusses possible links between working memory and subprocesses of composition, describes research designs and major methods for data elicitation in L1 research, and unveils the differential roles of working memory in L1 and L2 writing. The second section provides a synthesis of the methods and findings of existing research on working memory and L2 writing.

### *Working memory*

It is important to clarify the nature, architecture, and measurement of working memory because such knowledge is essential for an accurate understanding of the research on the associations between working memory and writing. Working memory is a cognitive system for simultaneous information manipulation, retention, and storage in ongoing tasks (Li, 2023). In order to have an accurate understanding of the architecture of working memory, it is necessary to start with Baddeley's (2015, 2017) model, which posits four components for working memory: the central executive coordinates the different components; the phonological loop stores and maintains auditory information; the visuospatial sketchpad is a storage space for visual and spatial information such as images, shapes, colors, and locations; and the episodic buffer is a transitional storage space between the two storage components and the central executive that integrates discrete information bits into larger units, links short-term and long-term memory, and binds information from different sources and information in different formats (e.g., auditory and visual information; colors and shapes).

There are two major models of working memory: the componential model and the unitary model, and advocates of the two models conceptualize and measure working memory in different ways. The componential model is championed by Baddeley and his colleagues. Based on this model, the components of working memory draw on different pools of resources and are independent of each other, and the storage components are proxies of working memory. In the research, the components are tested and investigated separately, and the primary focus of the research is on the role of the phonological loop in L1 vocabulary learning. The unitary model was initiated by Daneman and Carpenter (1980), who developed a reading span test that measures both the processing and storage components of working memory. In this model, working memory is a global construct that integrates the processing and storage components, which must be measured as a single concept, and there is a trade-off between the two components in that allocation of more resources to one leads to fewer resources for the other. This model emphasizes the importance of the central executive, and an extreme variant of the model holds that variation in working memory is primarily due to variation in attention control (Engle, 2002).

In line with the two theoretical models, working memory has been operationalized and measured differently in the research. Two broad categories of measures can be identified: simple and complex tasks, with the former tapping only the storage functions of working memory and the latter integrating both the storage and processing functions. In a typical simple working memory task, subjects are presented with lists of discrete, unrelated items in the auditory or visual mode and are asked to recall the items in oral or written form at the end of each list. In a complex working memory task, which measures both storage and processing functions, subjects are asked to perform two tasks—a primary memorization task and a secondary processing task. The central executive is measured differently in the two models. In the componential model, it is separate from the storage components and is fractionated into three functions that are measured accordingly (Miyake & Friedman, 2012). The three functions are inhibition, which refers to the ability to suppress irrelevant information; shifting, which refers to the ability to switch between different tasks; and updating, which refers to the ability to constantly monitor and update information in an ongoing task. In the unitary model, the central executive is measured using complex working memory tasks; in other words, it is equated with working memory. To unify the terminology, avoid confusion, and align with previous research, in this article, the phonological loop is named phonological short-term memory; verbal working memory refers to complex tasks in the verbal domain that measure both storage and processing functions; visual-spatial working memory refers to both simple and complex tasks used to measure the visuospatial sketchpad because evidence shows that the storage and processing functions of visual-spatial short-term memory are indistinguishable (Shah & Miyake, 1996); and the three executive functions are named separately using the three terms mentioned above: inhibition, shifting, and updating.

### ***Working memory and the processes of writing***

#### *Theoretical models*

Two major writing models have been drawn on in the research on working memory and writing: Hayes's model and Kellogg's model. Hayes's (Hayes, 1996; Hayes & Fowler, 1980) model identifies two major components: the task environment and the individual, which can be conveniently called learner-external and learner-internal dimensions of writing. The task environment refers to "all those factors influencing the writing task that lie outside of the writer's skin" (1996, p. 3), and these factors are further separated into two categories: a social component and a physical component. The social component includes the audience, the social environment, and the source text (as in integrated writing where the writer responds to a given text). The physical environment refers to the text produced so far and the writing medium (e.g., paper-based or screen-based writing). The individual dimension of writing includes learners' cognitive (working memory), conative (motivation), and affective (anxiety, self-efficacy, etc.) variation; the cognitive processes of writing such as planning, translation, transcription, and revision; and long-term memory, which refers to learners' previous knowledge about the topic, language, genre, etc. One striking aspect of Hayes' model is its emphasis on the importance of working memory in writing, which is evident in his claim that "all of the processes have access to working memory and carry out all nonautomated activities in working memory" (1996, p. 8).

Whereas Hayes's model concerns all aspects of writing, Kellogg's (1996; Kellogg et al., 2013) model focuses specifically on the role of working memory in writing. Unlike

Hayes who attaches importance to working memory in all phases of the writing process, Kellogg adopts a more nuanced approach, making claims about the differential roles of working memory in different steps or stages of writing. In this model, writing is divided into six processes: planning, translating, programming, executing, reading, and editing, which will be further detailed in later sections. Kellogg argued that the central executive, which refers to verbal working memory tapping into both the storage and processing functions of short-term memory, is involved in nearly all processes of writing except for transcribing or executing. Visual-spatial short-term memory is only important in planning, and the phonological loop or phonological short-term memory (the storage function) is only relevant to translating and reading. Similar to Hayes's model, Kellogg's model did not specify the roles of the three executive functions validated by Miyake and Friedman (2012)—inhibition, switching, and updating—a limitation acknowledged by Kellogg et al. (2013). The following sections elaborate on theoretical perspectives on the role of working memory in the subprocesses of writing, and the primary objective is to discuss the mechanism through which working memory is implicated in the writing process rather than synthesize empirical evidence. The discussion is primarily based on Kellogg's model, as this model informs most research on working memory and L2 writing.

### *Writing processes and working memory*

#### *Planning*

Planning involves idea generation and idea organization. Idea generation refers to the retrieval and selection of relevant information from long-term memory. Idea organization refers to the arrangement of ideas in logical order and imposition of a structure. Idea organization focuses primarily on the blueprint, overall structure, and framework, which may consist of the major sections of a paper and main ideas for each section, and may be represented as (1) an outline or cursive notes and (2) a mental sketch of the manuscript to be drafted. The mental sketch is more than an outline, which is restricted to the limited amount of information transcribed. Which types of working memory are involved in planning? In Kellogg's model, planning involves verbal working memory and visual-spatial working memory. The rationale for the involvement of verbal working memory is simple: writers must retrieve, select, organize, and retain information, and this process necessarily requires cognitive resources for information storage and processing.

Kellogg (1996) argued that planning is the only stage that involves the visuo-spatial sketchpad, whose main functions are to visualize ideas, organization, spatial layout of text, etc. Kellogg et al (2013) claimed that *idea generation* and *organization* are not distinguished in their model because the distinction does not make a difference for the role of working memory. In terms of *idea generation*, visual-spatial working memory may be conducive for planning concrete and physically tangible information or concepts, and this hypothesis has been confirmed by Kellogg et al. (2007), who found that visual working memory is only involved in concrete rather than abstract concept planning in a sentence-generation task. There is also evidence that visual-spatial working memory is drawn on in descriptive writing but not argumentative writing (Olive, 2022), further testifying to the involvement of visual-spatial working memory in compositions on concrete concepts and ideas, which are more likely to be involved in descriptive writing. However, the conclusions about planning were based on inferences because the studies did not examine planning directly. Regarding *idea organization*, the

relevance of visual-spatial working memory is premised on the speculation that writers create a mental sketch or projection of the text to be composed when conducting macro planning. Kellogg et al. (2013) suggested distinguishing visual and spatial working memory, and there is evidence to support this distinction. For example, Galbraith et al. (2005) reported that a spatial tracking task affected idea organization before text production but a visual noise task had no effect, suggesting the involvement of spatial rather than visual working memory in organization planning.

### *Translating*

The writer encodes the planned message verbally by retrieving and selecting linguistic items that match the content and arranging the items following morphosyntactic rules; this process is called translating. Translating involves grammatical encoding, phonological encoding, and orthographic encoding (Kellogg et al., 2013). To fully understand the translating process, it is useful to draw on Levelt's (1989) speech model, according to which in speech production, message planning is followed by formulation, which consists of two steps: grammatical encoding and phonological encoding. In writing, a third process orthographic encoding needs to be added because the orthographic forms of the linguistic items and the generated sentence must be visually represented. *Grammatical encoding* consists of procedures for accessing lemmas and procedures for syntactic building. Lemma has two components: concept and syntax. When a lemma is activated, the syntactic information of the lemma is made available, which calls for syntactic building procedures. The product of this stage is called surface structure, stored in the syntactic buffer. *Phonological encoding* involves retrieval of phonological information for lexical items and for the whole mental message. *Orthographic encoding*, which applies to writing but not speaking, refers to the activation of the spellings of words and the building of the visual image of the generated sentence.

To what extent is working memory involved in translating? First, phonological short-term memory is involved because the phonological aspects of the retrieved linguistic items may be activated and the composed sentence—the inner speech—may be held briefly in phonological short-term memory before it is executed (transcribed or typed out). Kellogg et al. (2007) argued that written language production may not involve the phonological loop, citing evidence that a patient with impaired phonological loop had normal speech and writing abilities and a patient with phonological impairments had no trouble retrieving orthographic forms. However, as Olive (2022) pointed out, although orthographic encoding may happen without phonological encoding, this may only be true of patients with impaired phonological short-term memory. Furthermore, even if phonological encoding is not required for written production, it may facilitate the retrieval of orthographic word forms, which in turn assists with grammatical encoding. Second, verbal working memory is involved because the writer must hold retrieved lemmas in an active state while performing syntactic processing and the major function of verbal working memory is simultaneous information storage and processing. Third, visual-spatial working memory is important for activating and retaining the orthographic representations of linguistic items and the generated sentence before it is transcribed. Kellogg (1996) did not posit a role for visual-spatial working memory in the translation stage, but this hypothesis may need to be modified.

### *Transcribing*

Transcribing refers to the process of transforming and transferring the mentally generated sentence to visible symbols through handwriting or typing. Transcribing has been operationalized as spelling, handwriting fluency, or style (punctuation and capitalization). Transcribing consists of programming and executing in Kellogg's (1996) model, which refer to formulating instructions or commands using code and implementing them, respectively. In Hayes's model (Hayes, 1996; Hayes & Flower, 1980), transcribing is part of translating rather than a separate process. Kellogg et al. (2013) predicted that verbal working memory is involved in programming and that spatial working memory may also be involved in execution on the grounds that "spatial parameters must be set in the motor programming of handwritten output ... [and] the spatial arrangement of the keyboard must be held in spatial WM [working memory] during the programming of the ballistic finger movements that strike the keys" (p. 176). There has been evidence that transcription speed and accuracy were indeed predicted by verbal working memory (Kim, 2022) and spatial working memory (Kellogg et al., 2013) as well as phonological short-term memory (Hayes & Chenoweth, 2006). It has also been argued that the role of working memory in transcribing is more evident for children than adults (Salas & Silvente, 2020) and for an unfamiliar transcription tool such as the computer than traditional handwriting stationary (Olive, 2022).

### *Reviewing*

Olive's (2022) recent review showed limited research on working memory and reviewing. Hayes (1996) argues that reviewing consists of reading and editing, both of which involve verbal working memory, according to Kellogg (1996). Reading may refer to reading a source text, which is relevant in integrated writing, or one's own text, which is relevant in independent writing. In the literature on working memory, the relationship between reading comprehension and working memory has been extensively researched (Li, 2023). Thus, the involvement of working memory in reading comprehension is less relevant in the context of writing, and what is of interest here is how working memory is associated with reading in integrated writing—such as in what way working memory is involved in the incorporation of information from the source text—or how it is related to understanding one's own text. Verbal working memory has been found to be a consistent predictor of reading comprehension, but the role phonological short-term memory in reading is inconsistent (Li, 2023). The empirical evidence is slightly deviant from Kellogg's (1996) model where phonological short-term memory and verbal working memory seem equally important in reading.

Editing refers to making changes to an existing text for improvement, and the changes may target linguistic errors, content, organization, typos, coherence, etc. Although editing has been assumed to target one's own writing, much of the research (e.g., Adams et al., 2010; Larigauderie et al., 2020) has examined editing given texts authored by others. As to the role of working memory, Kellogg (1996) claimed that only verbal working memory (called the central executive by Kellogg) is involved in editing but later recognized that phonological short-term memory may also be important (Kellogg et al., 2013). I argue that visual-spatial working memory is involved in editing when (1) keeping orthographic forms and a visual representation of the sentence in mind while making local changes and (2) keeping the locations of related information in mind so the current change is aligned with the big picture and previous sections while making macro changes that may affect other parts of the written text. One fruitful perspective for future research is examining the role of working memory in different

types of editing rather than, or in addition to, editing as a unitary or global construct. For example, fixing surface and local errors may just involve executive functions, whereas global and discourse-related errors may require both storage and processing.

### **Research design and data elicitation in L1 writing**

#### *Dual-task versus regression designs*

In L1 research, two major streams of research have been identified on the role of working memory in written composition: *the dual-task design* and *the regression (or correlational) design*. Early research, including the studies by Kellogg and Hayes and their colleagues (e.g., Chenoweth & Hayes, 2006; Kellogg et al., 2007), was conducted primarily using the dual-task method, which was borrowed from research on working memory (Baddeley et al., 1998). In this method, writers perform a concurrent, secondary working memory task while completing a writing task (the main task), and the purpose of the secondary task is to interfere with the writing process. There are two designs that represent different perspectives and lead to different methods of analysis. One is to examine whether *writing* is affected by the secondary working memory task, in which case writers may be given different working memory tasks (verbal, spatial, etc.) or no working memory task (control group who write performing a secondary task). The results will show whether writing is affected and/or whether it is affected differently by different secondary tasks, and then conclusions can be reached on whether working memory or different types of working memory are involved in writing. Another perspective is ascertaining whether writers' performance on *the secondary task* is affected by writing, and the purpose is to examine the cost in working memory resources that writing incurs. This approach is especially useful when learners' writing performance is not different between different conditions, and the idea is that although there is no difference in writing quality, the same quality of writing may have consumed different quantities of working memory resources, thus revealing the role of working memory. In this approach, it is preferable to have a control group who only perform the secondary working memory task, not the writing task, so as to determine whether participants' performance on the working memory task is affected by writing. Although a distinction was made between writing quality and working memory costs, researchers can combine the two approaches and investigate both in one experiment.

In existing research, it is unclear what the secondary task taps into or what kind of working memory this approach examines. Kellogg et al (2013) claimed that the purpose of the dual task is to pressure the central executive so learners have to multitask: "By requiring the individual to perform two tasks concurrently, the executive component of working memory is assessed as well as capacity of short-term memory stores" (p. 169). This quote seems to suggest that it is verbal working memory that is examined in the dual task. However, it can be argued that the tapped construct should be determined by the nature of the secondary task. For example, if the secondary task requires the writer to respond to meaningless stimuli such as unrelated digits, then it taps into phonological short-term memory. If the writer is required to perform a backward digit span task, which is a measure of verbal working memory, then the examined construct is verbal working memory. Furthermore, articulatory suppression, a method commonly used in dual-task research in writing where the writer repeats a syllable or word, is a method intended to disrupt the phonological loop in working memory research (Baddeley et al., 1998).

Whereas in the dual-task method writers perform two tasks simultaneously, in the correlational approach they perform the working memory task and the writing task separately. The correlational approach is also called the regression approach by Kellogg et al. (2013), and most studies synthesized in this article adopted the correlational approach. In this approach, writers take a test of working memory and complete a writing task, and then correlation-type analyses such as simple correlation, multiple regression, or structural equation modeling are performed to explore the associations between working memory and writing performance. In correlational research, a distinction can be made between direct and indirect approaches, with the former referring to examination of the direct effects of working memory on an outcome measure, and the latter to indirect effects working memory has on the outcome via another variable. For example, writers' variation in working memory may have a direct association with the quality of their writing because of limited working memory resources at their disposal during composition. It may have an indirect association with writing quality by influencing a factor or aspect that has a direct effect on writing quality, such as writers' previous linguistic knowledge (grammar, vocabulary, and pronunciation). Direct and indirect effects of working memory can be examined through path analysis, which is a variant of structural equation modeling.

Both dual-task and correlational approaches can be used to examine the relationship between working memory and various aspects of writing. For example, to examine the role of working memory and prewriting planning, in the dual task design, the researcher may ask writers to perform a secondary interference task while planning to see whether the outcome of planning is affected or whether planners' working memory performance is affected. In the correlational method, the researcher would measure writers' working memory separately and explore whether writers' working memory scores are correlated with the outcome of planning such as the number of words of planned notes.

#### *Process-base versus product-based approaches*

The research on working memory and writing can be divided into process- and product-based approaches. Process-based studies investigate how working memory or different components of working memory are implicated in the subprocesses of writing such as planning, translating, transcribing, and reviewing. Studies investigating the cognitive processes of writing via behaviors that occur during the writing process, such as pauses, repairs, and eye gazes, fall into this category (see Révész et al.'s and Torres's contributions to the special issue). Product-based studies examine the associations between working memory and the writing product evaluated holistically (overall writing performance) or analytically (aspects of writing performance; see Manchón et al.'s contribution to the special issue). Both approaches are needed and contribute to an accurate understanding of the cognitive dimensions of writing. A process-based approach facilitates our understanding of the mechanism through which working memory affects different processes of writing. A process-based approach entails a fine-grained, microscopic inspection and answers the questions of why working memory is or is not important and what processes or components are responsible for the effects of working memory or lack thereof. A product-based approach provides evidence on the importance of working memory for the outcome of writing, and it answers the question of whether, rather than how and why, it is important. The two approaches can be integrated to arrive at a more holistic understanding and excavate

the intricacy of the associations between the process and product of writing, in which case the design can be labeled a process-product approach.

Finally, the “process versus product” distinction is different from the “dual-task versus regression” distinction. The former distinction concerns whether the focus is on the writing process or the writing product, whereas the latter is based on whether writers have to perform concurrent writing and working memory tasks or complete them separately. Both process and product aspects of writing can be investigated by using a dual task or regression approach. The regression approach is alternatively called the correlational approach, but the term “correlational” is used mainly to refer to the research design rather than statistical analysis. In both distinctions, correlational analysis can be conducted in all research types if it fits the research goal or question.

### ***Differential roles of working memory in L1 and L2 writing***

Despite commonalities between L1 and L2 writing processes, there are differences between them that may lead to differential roles of working memory in the two types of writing. The different mechanisms between L1 and L2 writing justify research endeavors to unearth the role of working memory in L2 writing. The following is a list of dimensions along which working memory is likely involved in different ways in L1 and L2 writing; these may serve as fodder for thinking or hypotheses to be empirically verified.

- *Amount of WM investment.* L2 writing poses greater cognitive demands on the writer than L1 writing because of their insufficient linguistic and genre knowledge about writing in the target language; thus, working memory may play a greater role in L2 than L1 writing.
- *Allocation of WM resources.* Writers do more planning than translating when writing in their L1, but the reverse is true when writing in their L2 (Révész et al., 2017; Vallejos, 2020). Therefore, translating may consume more working memory resources than planning in L2 writing.
- *Translating.* First, in L1 writing, translating is automatic, whereas in L2 writing it is effortful and conscious. Therefore, translating consumes more working memory resources in L2 than L1 writing. Second, L1 writing may rely more on phonological short-term memory, whereas L2 writing may draw more on visuo-spatial working memory because of L2 writers’ (especially less proficient writers) heavier dependence on spelling or orthography. At a more advanced level where the L2 phonological system is more developed and automated, visuo-spatial working memory may become less important (Gunnarsson-Largy et al., 2019). Third, it is possible that the phonological form is always activated regardless of whether it is fully developed. In L2 writing, for example, a certain phonological form is activated, even though it is flawed, such as a surrogate form from the L1. Phonological short-term memory is therefore crucial.
- *Transcription.* Working memory may play a less important role in L1 transcription than L2 transcription. L2 learners, especially beginning learners, may not be familiar with the spelling or writing system and may therefore need to exert a substantial amount of cognitive resources in transcription. In a similar vein, working memory resources that are important for transcription may be less important at more advanced stages of L2 learning where learners have improved their transcription skills. Furthermore, working memory may play a greater role for learners of

languages whose writing systems are strikingly different from their native languages (e.g., L1 English and L2 Chinese or Arabic learners).

- *Editing.* L2 writers may focus more on surface or linguistic errors than content or discourse related errors. Thus, working memory is perhaps more relevant to the editing of language than content aspects of L2 writing.
- *L1 influence.* Inhibition may be especially important for L2 writing because L2 writers must inhibit genre and linguistic knowledge carried over from their L1, and the role of inhibition may be more evident in initial stages of L2 learners where L1 influence is greater compared with advanced stages. However, L1 influence is a complex issue (see Manchón & Polio, 2022), and its interaction with inhibition and other components of working memory needs nuanced theoretical elaboration.

### The research synthesis

The purpose of the following synthesis is to provide a review and critique of the research on working memory and L2 writing with a view to presenting the status quo, facilitating an accurate interpretation of existing research, and informing future research. The synthesis is guided by two broad research questions:

1. What methods have been used in the research examining the associations between working memory and L2 writing?
2. What has research demonstrated about the relationship between working memory and L2 writing processes and outcomes?

To identify the relevant literature, major databases in psychology, linguistics, and education were searched, including ERIC, LLBA, ProQuest Dissertations, PsychArticles, and PsycInfo. Key search words include terms relating to (1) working memory, components of working memory, and alternative terms and (2) writing and the processes of writing such as planning, translating, transcribing, and revision. Included in the synthesis were journal articles and doctoral dissertations. The methodological details and research findings of the retrieved research were recorded and coded. The coded methodological features include research foci (or research questions), sample characteristics, measurement of working memory, measurement of writing, methods of data elicitation for writing processes, and data analysis and reporting. Synthesis of the findings of the primary studies focuses on those relating to working memory, although some studies also examined other variables.

#### ***RQ 1: What methods have been used in the research?***

*Foci, research designs, and sample characteristics*

A total of 16 studies was retrieved that examined working memory and L2 writing processes and outcomes. Fourteen of the studies have been published in the past 5 years, suggesting that this is a new topic that has attracted interest only recently. Table 1 displays the main information about the synthesized studies including the research focus (or research questions), sample, type or component of working memory, outcome variable, and major findings. In terms of research focus, 12 out of the 16 studies are *correlational* studies examining the associations between working memory and overall writing performance or aspects of writing performance such as spelling (Arfé & Danzak 2020), pauses (Vallejos, 2020), and CAF (complexity, accuracy, and fluency; e.g.,

**Table 1.** Studies on the associations between working memory and L2 writing

Study	Foci	Sample	Working memory/test	Writing task/skill	Major findings
Arfé & Danzak, 2020	Predictors of L2 spelling	50 L1 Italian and L2 English 10 <sup>th</sup> and 11 <sup>th</sup> graders	<b>Inhibition</b> /Flanker task	Expository writing/L2 spelling	L2 proficiency was the strongest predictor of L2 spelling. <b>Inhibition</b> was predictive of two types of spelling errors
Cho, 2018	Task complexity, task modality, working memory, and L2 writing	39 university L1 Korean EFL learners	<b>Verbal working memory</b> /L2 reading span + operation span	Argumentative writing/CAF*	Task complexity increased complexity, reduced accuracy, and had no effect on fluency. Speaking was more accurate, less fluent, and more syntactically complex than writing. <b>Working memory had no effect.</b>
Gunnarsson-Largy et al., 2019	The role of working memory in L1 and L2 translation (formulation)	31 L1 French and 72 L2 French speakers	<b>Phonological short-term memory</b> /dual task: L2 nonword recall; <b>visual short-term memory</b> /dual task: grid recall;	Dictation /presence or absence of n' marking in French	During formulation (translation), L1 French writers relied more on <b>phonological short-term memory</b> and L2 French writers relied more on <b>visual short-term memory</b> but more advanced learners relied on <b>both phonological and visual short-term memory.</b>
Kim et al., 2021	Factors contributing to writing quality, text length, and writing fluency	100 ESL learners at a US university	<b>Verbal working memory</b> /L2 running span; <b>inhibition</b> / Stroop test	SAT writing; argumentative writing/overall ratings, fluency, text length	<b>Verbal working memory</b> predicted literacy (previous linguistic and world knowledge) which in turn predicted overall quality; <b>inhibition</b> predicted mean pause time; mean pause time predicted text quality.
Kormos & Sáfár, 2008	The relationship between working memory and writing proficiency	121 8 <sup>th</sup> grade EFL learners at a	<b>Verbal working memory</b> / backward digit span;	Cambridge First Certificate Exam/	<b>Verbal working memory</b> was not predictive of writing at a lower level of proficiency; it is

(Continued)

Table 1. (Continued)

Study	Foci	Sample	Working memory/test	Writing task/skill	Major findings
		bilingual school.	<b>phonological short-term memory</b> /nonword repetition	overall performance	unclear whether it's predictive of writing at a higher level. <b>Phonological short-term memory</b> was predictive of writing at the higher level of proficiency.
Leong et al., 2019	The role of working memory in different genres	129 middle school L2 Chinese learners in Hong Kong	<b>Verbal working memory</b> /L2 listening span tests	Narrative, expository, argumentative writing/overall writing performance	<b>Verbal working memory</b> was a significant predictor of explanation and near significant predictor of argumentation but nonsignificant predictor of narration.
Li & Roshan, 2019	The relationship between working memory and written corrective feedback	79 university EFL learners	<b>Verbal working memory</b> /reading span; <b>phonological short-term memory</b> /nonword repetition	Expository writing/English passive voice	<b>Verbal working memory</b> was a positive predictor of the effects of metalinguistic feedback, and <b>phonological short-term memory</b> was a negative predictor of the effects of direct corrective feedback plus revision.
Lu, 2010	Factors influencing L2 argumentative writing including L1 writing, genre knowledge, and working memory	136 Chinese EFL learners	<b>Verbal working memory</b> /L1 and L2 operation span tests	Argumentative writing/overall writing performance	<b>Verbal working memory</b> was not a predictor of overall writing; L2 knowledge was the strongest predictor of L2 writing, followed by genre knowledge and strategy use. L1 writing was not a predictor of L2 writing.
Mavrou, 2020	The associations between executive functions and writing performance	59 L2 Spanish learners	<b>Verbal working memory (named updating)</b> /operation span + running span; <b>visual-spatial working memory</b> /	Video narrative/CAF	<b>Verbal working memory</b> was predictive of syntactic complexity and accuracy. No other working memory

(Continued)

**Table 1.** (Continued)

Study	Foci	Sample	Working memory/test	Writing task/skill	Major findings
Michel et al., 2019	EFL learners' editing and writing performance at different grade levels and the role of working memory	94 grades 6 and 7 EFL learners at two primary schools in Hungary	backward Corsi block-tapping; <b>inhibition</b> /Stroop test; <b>shifting</b> /letter-digit switching <b>Composite working memory:</b> executive working memory/backward digit; phonological short-term memory/forward digit; spatial working memory/symmetry task.	Writing section of the TOEFL Junior test/overall writing ability, editing	measures were predictive of any CAF measures <b>Working memory</b> was only predictive of editing, not overall writing proficiency.
Peng et al., 2022	The role of phonological awareness, oral language, and working memory in L2 English writing.	374 grades 3, 4, and 5 Hispanic ELLs from three large school districts in the U.S.	<b>Composite working memory/</b> L2 and L1 conceptual span, listening span, rhyming, updating, visual matrix, mapping and directions.	Three subtests/overall writing ability using a latent factor represented by three aspects including writing a narrative	Both L1 Spanish <b>working memory</b> and L2 English working memory predicted writing; L2 oral language was a positive predictor but L1 oral language was a negative predictor; phonological awareness was a positive predictor in both languages.
Révész et al., 2017	The associations between writing behaviors and writing quality, and the predictive power of working memory on writing behaviors and quality.	30 ESL learners from a UK university.	<b>Verbal working memory/</b> operation span; <b>phonological short-term memory</b> /digit span, nonword repetition; <b>visual-spatial working memory/</b> ; <b>switching</b> /color shape, <b>visual-spatial short-term memory</b> /Corsi block, Corsi block backward; <b>inhibition:</b> stop signal.	IELTS writing; argumentative writing/overall writing, CAF	No significant correlations were found between <b>working memory</b> measures and overall writing proficiency. Regarding CAF, <b>phonological short-term memory</b> was a positive predictor of GSL1000 words (less lexical complexity); poor <b>shifting</b> predicted more use of GSL1000 and more logical connectors as opposed to causal, additive or contrastive connectors; better <b>shifting</b>

(Continued)

**Table 1.** (Continued)

Study	Foci	Sample	Working memory/test	Writing task/skill	Major findings
Vallejos, 2020	Cognitive processes of L1 and L2 writing, fluency, and working memory.	55 L2 Spanish learners at a US university	<b>Verbal working memory/</b> operation span, switching/ color shape	Argumentative writing task/pauses	correlated with shorter pauses; better <b>verbal working memory</b> associated with less frequent inter-paragraph pauses; better <b>visual short-term memory</b> associated with less frequent eye-gazes at instructions during pauses. In L1 writing, writers planned more than translated; in L2 writing, they translated more than planned. <b>Greater working memory led to an increase in the number of pauses between sentences.</b>
Vasylet & Marin, 2020	Working memory's associations with writing performance at different proficiency levels	59 L1 Spanish/ Catalan L2 English sophomores at a Spanish university.	<b>Verbal working memory/</b> reading span	Narrative video retelling/overall writing and CAF	<b>Verbal working memory</b> was not a predictor of holistic scores of video narratives; it was predictive of low proficiency learners' accuracy and high proficiency learners' lexical sophistication.
Zabihi, 2018	Direct and indirect effects of working memory, anxiety, and self-efficacy on writing complexity, accuracy, and fluency	232 adult EFL students	<b>Verbal working memory/</b> operation span	Picture narrative/CAF	Verbal working memory was a significant predictor of fluency and syntactic complexity but a negative predictor of accuracy. Both anxiety and self-efficacy predicted CAF; self-efficacy also indirectly predicted CAF via anxiety.
Zalbidea, 2017	Effects of task complexity, task modality, working memory, and writing performance	32 L1 English undergraduate students in a 3rd semester L2 Spanish class.	<b>Verbal working memory/</b> operation span	Argumentative writing task/CAF	<b>Verbal working memory</b> predicted accuracy in complex writing tasks and syntactic complexity (use of subordinating conjunctions) in speaking tasks.

Note: CAF: complexity, accuracy, and fluency.

Vasylet & Marín, 2020). These studies did not involve variable manipulation, and in some studies working memory was examined as a predictor of writing performance together with other predictors such as previous L2 knowledge (Lu, 2010), oral language (Peng et al., 2022), or anxiety (Zabihi, 2018). Among the 12 correlational studies, Kormos and Sáfár (2008) and Vasylet and Marín (2020) examined whether the role of working memory was moderated by learners' L2 proficiency; Leong et al. (2019) investigated whether the effects of working memory varied as a function of genre. The remaining four studies are *experimental* studies that involved systematic variable manipulation. One of the four studies examined the roles of different types of working memory in translating (referred to as formulation by the authors) at different levels of proficiency (Gunnarsson-Largy et al., 2019); one investigated the associations between working memory and different types of written corrective feedback (Li & Roshan, 2019); and two studies examined the interface between task complexity, task modality, and working memory (Cho, 2018; Zalbidea, 2017).

Among the 16 L2 studies, only Gunnarsson-Largy et al. (2019) used the dual-task approach, and the remaining studies used the regression approach. In this dual-task study, participants performed a dictation task while performing a phonological (memorizing three nonwords) or visual (memorizing grid squares) concurrent task intended to interfere with phonological short-term memory or visual-spatial short-term memory. The data were analyzed in three ways: (1) comparing learners' performances on the dictation tasks to see whether the secondary working memory tasks affected learners' dictation performance, (2) comparing learners' performances on the secondary working memory tasks to evaluate the cognitive costs of the main tasks, and (3) performing correlations between secondary and main task performances within the same group to determine whether there was a trade-off between the two tasks or whether learners were engaged in the main task.

In studies using a regression approach, writers' working memory and writing ability were tested separately and statistical analyses were performed to determine whether the two scores correlated. Three design features of the synthesized studies should be highlighted. One is the examination of indirect effects or mediated effects of working memory (or another predictor) on outcome measures, which means that working memory has a direct effect on another variable, which in turn affects writing performance. The mediating approach was used in Kim et al. (2021), which examined working memory's indirect contribution to writing performance via literacy skills, and in Zabihi (2018), which investigated whether working memory, self-efficacy, and anxiety were directly predictive of writing performance and whether self-efficacy was also indirectly predictive of writing performance via anxiety. A second design feature is the focus on the moderated effects of working memory on writing outcomes; that is, the role of working memory depends on a third factor, such as type of corrective feedback (Li & Roshan, 2019), genre (Leong et al., 2019), learner proficiency (Arfé & Danzak, 2020; Kormos & Sáfár, 2008), and task complexity (Cho, 2018; Zalbidea, 2017). The third design feature is the examination of latent or composite variables, which is typical of large-scale studies using multiple measures for the same constructs to identify the relationships between multiple variables, such as Kim et al. (2021), Lu (2010), Mavrou (2020), and Peng et al. (2022). A latent variable is the underlying trait, ability, or skill represented or indexed by multiple observable behaviors or phenomena. For example, Peng et al. administered six measures of working memory, three measures of writing competence, two measures of oral language, and five measures of phonological awareness, and the researchers tested the relationships between the latent variables represented by the concrete measures.

Regarding the distinction between process- and product-based studies, only four out of the 16 studies examined writing processes (translating, editing) and writing behaviors (repairs, pauses, and eye gazes); other studies examined the product of writing such as overall writing quality or specific aspects of writing quality such as complexity, accuracy, and fluency.

Information regarding sample characteristics and research contexts is as follows. Among the 16 studies, 11 were conducted with university students, one high school students, two middle school students, and two elementary school students. Ten studies were conducted in foreign language settings where the target language was not spoken outside of class, and six were conducted in second language settings where the target language was used both inside and outside of class. The target languages of the 15 studies were relatively homogeneous: in 11 studies, English was the target language; three studies targeted Spanish as the L2; Chinese and French were each examined in one study. The L1s of the participants in this data set were varied, including Chinese, English, Korean, Italian, Hungarian, Persian, French, Spanish, or mixed in the case of a sample consisting of international students with varied L1 backgrounds.

### *Measurement of working memory*

Verbal working memory (both storage and processing functions) was investigated in 14 out of the 16 synthesized studies, phonological short-term memory in five studies, and visual-spatial working memory in four studies. The three executive functions of working memory were examined as follows: inhibition in three studies, shifting/switching in two studies, and updating in one study. Verbal working memory was measured by using operation span tests (where learners are asked to judge the correctness of math equations and remember unrelated symbols) in seven studies; reading span tests in three studies; backward digit tests in two studies; and listening span, conceptual span, and rhyming tests each in one study. Phonological short-term memory was measured via nonword recall in four studies and forward digit span in one study. Visual-spatial working memory was gauged by means of symmetry tasks, visual matrix, mapping and directions, Corsi block, and Corsi block backward. The measures of verbal working memory and phonological short-term memory reported above can be divided into two major categories based on whether the stimuli are verbal or nonverbal. Thus, measures based on math equations (operation span) and digits (forward and backward digit) are nonverbal, whereas measures based on linguistic stimuli such as reading span, listening span, word span, and nonword span are nonverbal tests. The choice between verbal and nonverbal tests will be revisited in the discussion section.

As to the three executive functions, inhibition was tested via Flanker, Stroop, and stop signal tasks; shifting was measured through a letter-digit switching task; and updating was gauged by means of running memory (n-back) tests. One prominent practice of the primary studies is to use a composite or factor score based on exploratory factor analysis or confirmatory factor analysis (in structural equation modeling) that represents the conglomerate construct of working memory. For example, Mavrou (2020) combined operation span and running memory and labeled the variable “updating.” Michel et al. (2019) created an overarching variable of working memory comprising measures of verbal working memory, phonological short-term memory, and visual-spatial working memory. Peng et al. (2022) identified a common factor underlying multiple measures of verbal working memory and visual-spatial working memory. There was also confusion over the matching between measures and

constructs; for example, the updating function of the central executive was found to load onto the same factor as verbal working memory (Mavrou, 2020; Peng et al., 2022), which casts doubt on whether updating is a measure of verbal working memory or an executive function.

Two aspects of the measurement of working memory that may have affected the results of the primary research are the methods of scoring and the language of the stimuli (i.e., L1 or L2). Span tests—tests where items are presented in groups of different numbers of items—are the most typical tests of working memory. Span tests can be scored in two ways: span-based and item-based. In span-based scoring, the test score is based on the longest set that the participant can recall, which yields a small range, such as 2–6. In item-based scoring, the test is scored based on all correctly recalled items, which gives a larger range and is desirable in individual-difference research. In this data set, out of the 20 related cases, 14 used item-based scoring and six followed span-based scoring. In item-based scoring, two methodological features that may affect the results are whether the order of recalled items is scored and whether the processing components such as veracity judgement and reaction time are scored. Among the primary studies, only Kim et al. (2021) mentioned that the order of recalled items was not scored and only Li and Roshan (2019) scored both the recall and processing (reaction time and judgements of math equations) components. In terms of the language of stimuli, in five out of the 16 studies, learners' L2 was used to create test items.

#### *Writing tasks and measures*

The measurement of the dependent variable of the research is crucial for the validity of the findings. The methodological dimension was coded as eliciting tasks, measured constructs, operationalization, and scoring of writing performance. Eliciting tasks refer to writing prompts used to elicit learners' writing samples. Seven out of the 16 studies used argumentative writing tasks, five used narrative writing tasks, three used expository writing tasks, one used a dictation task to examine the translating process of writing, and two used multiple genres. One study employed an integrated writing task (listen and write), and all other studies used independent writing tasks. Six studies used test prompts from standardized proficiency tests such as TOEFL (in three studies), IELTS, SAT, and Cambridge First Certificate Exam. The writing tasks in the primary studies were implemented differently in terms of writing time, availability of planning time, and word limit. Only eight studies reported the time writers were allowed to complete the writing task, ranging from 10 to 45 min, and among the studies reporting writing time, five allowed more than 20 min. Only four studies reported the number of words writers were expected to write: two set the limit as 200, one as 250, and one as 50–150 Chinese characters. Only two studies reported allowing planning time (3 and 5 min, respectively) before writing commenced. These methodological differences pose different task demands, and it is unclear to what extent the obtained results were due to task demands, which could be possible extraneous variables. Also, regarding justifications for the methodological decisions, only one study validated the time limit by piloting the task with a small number of learners before the main task. One of the two studies on task complexity reported independent evidence for the validity of the construct of task complexity.

Regarding the measured construct, seven studies assessed overall writing performance; six examined complexity, accuracy, and fluency (CAF) together; one investigated spelling; two focused on accuracy in using specific structures; and one study investigated other specific measures as dependent variables such as p-bursts and

pauses. The measured constructs were operationalized and scored in various ways. Overall ratings are proxies of global writing ability assessed through human judgments. The rated aspects included in the rubrics, however, varied significantly including different configurations of content, organization, language use (accuracy, grammar, lexical variety), coherence, topic development, syntactic variety, and so on. Among the seven studies using overall ratings, four had two raters evaluate each writing sample, one had one rater, one had three raters, and one did not report the number of raters. Five studies stated that the raters received training without providing details on the training. The raters were PhD students, teachers, and IELTS trainers. In four studies, the ratings by different raters were averaged, and in one study the ratings were agreed upon by the two raters. The ratings seemed analytic—namely, that different aspects of writing were rated, but scores were holistic in that composite scores rather than discrete scores representing different aspects of writing were analyzed.

The use of CAF measures to assess writing is a general trend in the retrieved studies. Complexity can be divided into syntactic, lexical, discourse, and propositional complexity. Syntactic complexity was indexed by length-based measures such as mean length of T-unit/clause; subordination—namely, the use of subordinate clauses; coordination (use of coordinate sentences); and nominal (noun-related) constructions such as nouns with pre- or postmodifiers, nominal clauses, and gerunds and infinitives in subject position (Cho, 2018). Lexical complexity can be further divided into lexical variety—use of different words—and lexical sophistication—use of less frequent words. Lexical variety was measured via type-token ratio and D, and lexical sophistication through Advanced Guiraud. Discourse complexity was assessed via cohesive devices including causal (e.g., “because”), logical (e.g., “therefore”), additive (e.g., “and”), or contrastive (e.g., “however”). Propositional complexity was operationalized as the idea unit, which refers to “a meaningful, semantically integral chunk of discourse” (Vasylet & Marín, 2020, p. 5). Accuracy was measured via error-based indices such as the number of error-free clauses or T-units, number of errors per hundred words, etc. Fluency was operationalized as (1) speed such as number of words per minute or total number of words/T-units/clauses (when the writing time was the same for all participants) and (2) pause-related indices such as number of pauses per 100 hundreds and mean pause time. In the primary studies, a pause typically referred to an interval longer than 0.2 s between two writing bursts. It is noteworthy that there was confusion over what construct a measure represents. For example, mean length of T-unit was considered a measure of syntactic complexity in Zalbidea (2017) but a measure of fluency in Zabihi (2018). Guiraud’s index was considered a proxy of lexical variety by Zalbidea (2017) but a measure of lexical sophistication by Vasylet and Marín (2021).

#### *Methods of data elicitation for writing processes*

It is important to know how the processes of writing were examined, although most L2 studies examined the product or outcomes of writing. In the few studies that examined working memory’s associations with the cognitive processes of writing, four tools were used: keystroke logging, stimulated recall, questionnaires, and eye tracking (Cho, 2018; Kim et al., 2021, Révész et al., 2017, Vallejos, 2020). Keystroke logging records online writing behaviors such as pauses and repairs, which can be directly analyzed or serve as prompts for stimulated recall. All studies involving keystroke logging used the free software program InputLog (<https://www.inputlog.net/>; Leijten & Van Waes, 2013). Stimulated recall is conducted after a writing task was completed, and during the recall, writers were asked to report what they were thinking at a given point, such as during

pauses, or why they made repairs. The recalls were then classified according to writing processes such as planning, translating, and monitoring (Révész et al., 2017; Vallejos, 2020). Another tool used to investigate writing processes is the questionnaire where writers were asked to respond to Likert-type scale questions on their cognitive behaviors before or during writing. Obviously, the questions were prepared by the researcher a priori and imposed on the respondent. Lu (2010) used a questionnaire as a tool to measure writing strategies rather than writing processes, but many questions in the questionnaire were concerned with writers' psychological processes during writing (e.g., "Before writing, I thought about the structure of the paper"). Another technique used in the research is eye tracking, which records the frequency and duration of writers' eye gazes during writing such as while they are pausing. For example, Révész et al. (2017; this special issue) tracked L2 writers' eye gazes during pauses and found that writers with smaller working memory capacities viewed writing instructions more frequently than did those with larger working memory capacities.

### *Data analysis and reporting*

Factor analysis was used in five studies, multiple regression in seven studies, and structural equation modeling analysis in three studies. In four studies, factor analysis was used to reduce the number of independent variables and map the relationships between measures and the underlying constructs, followed by multiple regression analysis where factor scores served as predictors. In three studies, the role of working memory was examined through simple correlation analysis. In two studies, measures of different types of working memory such as verbal working memory, visuospatial working memory, and updating function of the central executive were combined and treated as one variable. With respect to reporting practices, all studies reported descriptive statistics such as means and standard deviations or correlation coefficients, six studies did not report checking assumptions for inferential statistical analyses, four studies failed to report any reliability indices, six reported reliability indices for certain but not all tests, and no study performed a power analysis. Furthermore, two studies reported Cohen's *d* as effect sizes, and one study (Peng et al., 2022) interpreted the magnitudes of regression coefficients, which is unusual in L2 research where sizes or weights of regression coefficients and factor loadings are mostly ignored.

### **RQ 2: What has been found about the role of working memory in L2 writing?**

In this section, the results of the 16 studies, which are summarized in Table 1, are organized by components of working memory in focus—the independent variables of the primary studies—and their associations with outcome variables. Types of working memory include verbal working memory (storage and processing of verbal information), phonological short-term memory (only storage), visual-spatial working memory (mix of measures of storage and processing of visual-spatial information), and executive functions. The two studies investigating composite working memory are included in the category of verbal working memory, although the construct also consists of other working memory measures. Within the section on a particular component of working memory, results are categorized by other independent variables such as task complexity or prominent methodological features such as outcome measures. Due to limited research on phonological short-term memory and visual-spatial working memory, results on the two types of working memory are reported in one section. Subsections

under each working memory component are structured based on outcome measures such as overall writing quality and CAF (complexity, accuracy, and fluency).

### *Verbal working memory*

Verbal working memory refers to the ability to store and process verbal information simultaneously. Studies in this category primarily examined the associations between verbal working memory and two broad categories of outcome measures: overall writing quality and CAF (complexity, accuracy, and fluency) dimensions. Overall quality was typically rated by two experts based on impressionistic judgements, and CAF measures were based on script or textual analyses. A total of eight studies examined the associations between verbal working memory and overall L2 writing quality: Kim et al. (2021), Kormos and Sáfár (2008), Leong et al. (2019), Lu (2010), Michel et al. (2019), Peng et al. (2022), Révész et al. (2017), and Vasylet and Marín (2020). Six of the eight studies showed no significant correlations between verbal working memory and overall writing quality. Of the remaining two studies examining overall writing quality, Leong et al. (2019) found verbal working memory a significant predictor of expository writing, a near-significant predictor of argumentative writing, but a nonsignificant predictor of narrative writing. Peng et al. (2022) reported that a latent working memory factor underlying measures of verbal and visual-spatial working memory was predictive of a latent writing factor represented by three measures: a language test, a test of mechanics, and a narrative writing task. These studies showed that in general verbal working memory is not a significant predictor of overall writing performance. However, it may have differential effects on different genres, and its effects may become evident when other components of working memory are involved and when writing is measured more globally through multiple measures.

The seven studies examining CAF as dependent variables showed the following findings. The two studies on task complexity obtained different results: Cho (2018) failed to find any effects for verbal working memory, whereas Zalbidea (2017) showed that verbal working memory predicted accuracy in gender agreement (better memory was related to fewer errors) but not number agreement in a complex writing task and that it was not predictive of writing in simple tasks. Li and Roshan's (2019) study demonstrated that verbal working memory was a positive predictor of the effectiveness of metalinguistic feedback but not other feedback types such as direct correction. Vasylets and Marín (2020) reported that verbal working memory was predictive of low-proficiency learners' accuracy and high-proficiency learners' lexical sophistication. Mavrou (2020) found verbal working memory predictive of syntactic complexity and accuracy. Zabihi (2018) found that verbal working memory was a significant predictor of fluency and syntactic complexity but a negative predictor of accuracy. Révész et al. (2017) demonstrated that better verbal working memory was associated with less frequent interparagraph pauses. Vallejos (2020) showed that better verbal working memory was predictive of more frequent between-sentence pauses.

To summarize, these studies seem to show the following results regarding CAF (also see Kormos and Manchón et al. in this special issue). For accuracy, the role of verbal working memory is constrained by task complexity, feedback type, and proficiency level; working memory may have a negative effect on accuracy. For complexity, verbal working memory was a positive predictor of syntactic complexity measured as subordination, and it was correlated with high-level learners' lexical complexity. For fluency, verbal working memory may lead to less frequent interparagraph pauses but more frequent between-sentence pauses.

### *Phonological short-term memory and visual-spatial working memory*

Only a few studies examined the role of phonological short-term memory in L2 writing. Kormos and Sáfár (2008) found a significant correlation between this component of working memory and overall writing performance at a higher proficiency level, but no significant correlation was found at a lower level. Using a dual-task approach where writers' working memory is interfered, Gunnarsson-Largy et al. (2019) found that during translation (called formulation by authors), L1 French writers relied more on phonological short-term memory than L2 French writers and that more advanced L2 learners relied more on phonological short-term memory than beginners. Li and Roshan (2019) showed that phonological short-term memory was a negative predictor of the effects of direct corrective feedback plus revision. Révész et al. (2017) found a strong correlation between phonological short-term memory and L2 writers' use of words from the most frequent 1,000 words, and the researchers interpreted this finding as showing a negative role for phonological short-term memory in lexical complexity.

As to visual-spatial working memory, Gunnarsson-Largy et al. (2019) found that L2 writers relied more on visual-spatial working memory than L1 writers in translation and that visual-spatial working memory was less important for more advanced L2 writers. Mavrou (2020) found no significant effects for visual-spatial working memory on CAF measures. Révész et al. (2017) showed that L2 writers with lower visual-spatial capacities gazed at writing instructions more frequently.

### *Executive functions*

The three executive functions of working memory—inhibition, shifting (switching), and updating—have received limited attention in the research. Arfé and Danzak (2020) found that inhibition was a positive predictor of spelling accuracy in expository writing, but the effect was found only for morphological and code-switching errors, not phonological and orthographic errors. The authors attributed the result to transfer of L1 features to L2 writing. Kim et al. (2021) found inhibition a positive predictor of pause length—namely, that writers who were better at inhibiting irrelevant information paused for a shorter time. Two studies were conducted on shifting. Mavrou (2020) did not find shifting to be a predictor of CAF measures, and neither did Vallejos (2020). Révész et al. (2017) found that participants who had weaker shifting abilities used more words from the most frequent 1,000 words, used fewer logical connectors, and paused for longer periods between sentences. Kim et al. (2021) is the only study that examined the updating function of working memory independently, although it was referred to as working memory. The study failed to find a significant effect for updating on writing quality, although it was predictive of literacy skills (L2 vocabulary, reading comprehension, and general world knowledge), which in turn predicted writing quality. However, the indirect effect of working memory on writing quality wasn't significant.

## **Discussion**

The discussion is structured around methodological issues and research findings. The methodological discussion centers on sampling, measurement of working memory, measurement of writing performance, methods of data elicitation, and analysis and reporting. Methodological issues are discussed by presenting the status quo, identifying pitfalls, recommending solutions, and suggesting directions. The findings on the role of working memory in L2 writing are interpreted by resorting to theories, consulting

research methods, identifying patterns, and resolving disparities. The interpretations are accompanied and followed by a checklist of variables that may moderate or mediate the effects of working memory on L2 writing processes and outcomes, with a view to informing future researchers of possible items to be placed on their research agenda.

### *Methodological issues*

#### *Sampling*

Several issues related to sampling merit researchers' attention. First, sampling bears on results, and therefore sample characteristics such as age, learning experience, learning stage, and proficiency level should be considered when making sampling decisions and consulted when results are interpreted. For example, working memory is more likely to be drawn on in transcription in initial L2 learning when learners are less familiar with the L2 orthographic system. Second, one issue that emerged from the synthesized research is biased sampling, which refers to the possibility that the selected sample is unrepresentative of the whole learner population. One of the synthesized studies, for example, was conducted with elite EFL learners at a prestigious university with stringent admission criteria. The truncated sample may have been partly responsible for the lack of significant results—namely, the lack of variation in the sample's working memory and their writing ability. For example, in this particular study, the mean score was 92% for L1 working memory and 85% for L2 working memory. Third, sampling heterogeneity may confound the examined variable, and the obtained results may be due to unexamined variables—a limitation that can be minimized by measuring and analyzing potential confounding variables as random factors using mixed-effects model analysis. Finally, decisions on sample size should be based on a power analysis, but no study in this data set conducted a power analysis or justified the sample size. Furthermore, an a posteriori power analysis can be conducted to determine the power of the study given the sample size and obtained results.

#### *Measurement of working memory*

To start with, there has been conceptual and methodological confusion over working memory. For instance, the operation span test is a measure of verbal working memory but was considered a measure of the updating or shifting function of the central executive in a study of this data set. Second, L2 working memory tests are subject to learners' L2 proficiency and should therefore be avoided; test stimuli should be based on learners' L1 or language neutral. However, five out of the 16 primary studies used L2 working memory tests. Third, a decision to treat working memory as a global construct or a componential construct measured separately concerns whether to fuse different types of working memory and use a composite score or to examine different types of working memory using discrete scores in data analysis. Michel et al. (2019) and Peng et al. (2022) combined three different types of working memory—executive working memory, phonological short-term memory, and spatial working memory—and used a composite score in data analysis. Most other studies, however, examined the unique influence of different types of working memory. Whether to treat working memory as a latent factor or a componential construct whose components contribute uniquely to L2 writing is a decision to be made based on theory and evidence. Fourth, the data set of this synthesis shows a predominance of operation span (judging the veracity of math equations followed by letter recall) as a measure of verbal working memory, and digit span tests were frequently used as a measure of phonological short-term memory.

However, verbal tests should be prioritized over nonverbal tests because of the alignment between verbal measures and language learning, and there has been robust evidence for the stronger predictive power of verbal tests on outcomes of language learning (Li, 2017a; Wen & Li, 2019). Fifth, it is necessary to include the processing components (reaction time and plausibility judgement) of verbal working memory because they make a difference in results (Sagarra, 2017). Alternatively, the researcher may run analyses with and without processing components before making a decision. Sixth, there may be a need to examine visual and spatial working memory separately in light of evidence for their different roles in writing processes (Kellogg et al., 2013).

I make the following recommendations to address the above and other issues identified in the methodological synthesis based on the literature and best practices (see Li, 2023, for further information on the methods of working memory):

- Justify the decision to treat working memory as a latent (global) or componential construct theoretically and empirically.
- Use verbal rather than nonverbal tests, and for nonword recall tests, the stimuli should be based on an unfamiliar rather than a familiar language (e.g., the L2).
- Use L1 rather than L2 tests.
- Use item-based (scoring all items) rather than span-based (scoring the maximum number of items test takers can memorize) scoring.
- Include processing components (represented by reaction times and plausibility judgments) of working memory in scoring or exclude them after making sure that they have no influence on the results.
- Examine visual and spatial working memory separately.

### **Measurement of writing performance**

The measurement of writing performance, which is the dependent variable of the research, is another methodological aspect that affects the results and is therefore essential for the validity of the findings. The measurement of writing performance can be divided into three components: the writing task, the implementation of the writing task, and scoring. In terms of writing tasks, argumentative writing was the most frequently used writing type, followed by narrative writing, and expository writing was the least frequent. Most studies do not justify their choice of writing tasks, but the rationales behind the researchers' decisions can be inferred. The popularity of argumentative writing is probably because "the ability to produce a well argued essay is crucial in academic contexts in higher education" (Kim et al., 2021, p. 5), and L2 writing research typically targets learners who study an L2 for academic purposes. A number of studies borrowed argumentative writing prompts from standardized proficiency tests such as TOEFL (Michel et al., 2019) and IELTS (Révész et al., 2017). Narrative writing is close to learners' daily lives and is one of "the most universal types of discourse in everyday language production" (Vasylet & Marín, 2020, p. 5), which may explain why it's a common writing genre in L2 research. Clearly, the choice of writing tasks is mostly based on pedagogical considerations, but researchers must also consider what effect a certain writing task may have on writing processes and outcomes due to the type and amount of cognitive demand it imposes on the writer. Therefore, task selection needs to be justified theoretically and empirically and task demands should be consulted when a study is designed and when results are interpreted. There is evidence for this recommendation, for example genre (e.g., Leong et al., 2019) and task complexity (Manchón

et al., this special issue; Zalbidea, 2017) have been found to moderate the associations between working memory and L2 writing performance.

Whereas the above discussion concerns the selection of a writing task, the way a selected task is implemented or the procedural aspects of a writing task are equally important. There was a high degree of heterogeneity in the way writing tasks were implemented in terms of availability of planning time before writing, time limit, word limit, etc. Similar to task selection, decisions on task implementation were rarely justified or mentioned in the reviewed studies, and yet methodological variation may affect task demands, which may in turn cause differences in the role of working memory in writing. For example, research on L2 speech production shows that working memory was implicated in unpressured speech but not in pressured speech (Li & Fu, 2018). Planning type may also influence the role of working memory in L2 writing, despite a lack of research.

Regarding scoring, two major methods can be identified in the data set: rating of overall writing proficiency and CAF (complexity, accuracy, and fluency); the former is subjective and the latter objective, which is probably why CAF measures have become increasingly popular in L2 writing (and speaking) research. However, ratings may capture aspects that CAF measures cannot such as content and organization. Therefore, these two scoring methods are complementary and should be used together to obtain a more holistic picture of the results. Overall writing proficiency was rated in many different ways, but the construct validity of the measures needs to be theoretically clarified and empirically tested. One example of theory-based measurement of writing performance is Leong et al. (2019), where the rating rubric was based on Halliday's systematic function grammar. Despite the increasing use of CAF measures, there has been confusion over what the measures represent. For example, mean length of T-unit was considered a measure of syntactic complexity in Zalbidea (2017) but a measure of fluency in Zabihi (2018).

Based on the methodological synthesis, I recommend a nuanced approach to scoring writing performance, which can be interpreted as follows. First, regarding subjective measures such as overall rating, discrete scores representing aspects of writing proficiency such as language and content rather than composite scores should be used in data analysis to explore whether working memory is associated with specific aspects of writing. However, Lu (2010) found that content and language scores were highly correlated and were therefore combined in data analysis. Thus, it is possible that content and language are not distinguishable in overall ratings, but there needs to be evidence for this conclusion, and more nuanced rubrics may make a difference in the results. Second, a nuanced approach also applies to objective measures such as complexity, accuracy, and fluency. For example, Zalbidea (2017) found that verbal working memory measured by an operation span test predicted the accuracy of gender agreement but not number agreement. For fluency, greater updating ability was found to be associated with shorter pauses between paragraphs in Révész et al. (2017) but with more frequent pauses between sentences in Vallejos (2020). Third, a nuanced approach requires genre-specific measurement tasks and scoring methods that are more representative of the discourse and linguistic features of narrative, argumentative, and expository writing and other genres. Fourth, a nuanced approach also requires the use of task-specific measures that represent the kind of ability a certain writing task purports to measure such as integrated writing (listen and write, read and write, etc.), which should be evaluated differently from independent writing because of the different purposes and demands of the two types of writing tasks.

### *Data elicitation*

Most L2 studies in this data set examined the product aspects of writing, and only a few studies probed into the processes of writing (Lu, 2010; Révész et al., 2017; Vallejos, 2020). Common methods for capturing writing processes can be divided into two categories: online and offline methods. Online methods are used during writing to capture covert and overt writing behaviors, whereas offline methods are used after the writing task is completed. Online methods include think-aloud protocol, where the writer is stopped intermittently and asked to reflect on current mental behaviors, and cued recall, where writers are trained on certain processes such as planning, translation, and review and are asked to report which process they are engaged on the fly. For example, in Kellogg's (e.g., 1987) research, writers heard a beep every 15–45 s during writing, said "stop" as quickly as possible, and then reported whether their thoughts best reflected planning, translating, reviewing, or a process unrelated to the three. One major offline method is stimulated recall, where writers reflect on what they were thinking during writing or during pauses. Stimulated recall can be aided by keystroke-logging software that captures the writing process. In L2 research on working memory and writing, stimulated recall and keystroke logging have been used (Révész et al., 2017) but verbal protocol and online cued recall have not (but see Torres, this special issue). What is encouraging is that Révész et al. and Vallejos (2020) coded writers' stimulated recalls following Kellogg's (1996) model, demonstrating that data coding was based on a theoretical model, thus enhancing construct validity. Révész et al. also used eye tracking to record writers' eye gazes.

Next, I discuss how the subprocesses of writing such as planning, translating, transcribing, and editing can be and/or have been examined. Although some of the subprocesses have not been examined in L2 research, I hope to draw attention to the relevant methods and inspire more research. Planning may happen before or during writing, and prewriting planning can be examined as planned notes or reflections on the planning process during planning via think-aloud protocol or after planning through stimulated recall. Translation has been investigated via oral language production (e.g., telling a narrative) based on the argument that "preverbal ideas and thoughts have to be encoded into oral language before being transcribed into written texts" (Kim & Schatschneider, 2017, p. 36), writing bursts (segments between pauses; Kim, 2022), sentence writing (Levy & Marek, 1999), or simply the first draft of a composition (Vandenberg & Swanson, 2007). In L2 research, Gunnarsson-Largy et al. (2019) operationalized translation as a dictation task, and other studies used stimulated recall (Révész et al., 2017). Transcription has not been investigated in L2 research. In L1 research, transcription has been investigated separately from writing and operationalized as handwriting fluency (Salas & Silvente, 2020), copying a text from one computer to another (Hayes & Chenoweth, 2006), spelling (Kim & Schatschneider, 2017), and so on. Editing was examined by one L2 study and was operationalized as proofreading where learners were asked to correct errors built in two paragraphs (Michel et al., 2019). In L1 research, it has been examined mainly through proofreading a given text (e.g., Larigauderie et al., 2020), and other tasks that have been used include rewriting illogical sentences (Swanson & Berninger, 1996) or simply a revised draft of one's own composition (Vandenberg & Swanson, 2007).

### *Analysis and reporting*

It is advisable to make justifiable or informed decisions in data analysis. For example, Mavrou (2020) performed an exploratory factor analysis on the measures of working

memory and the measures of CAF. The factor analysis showed that the three measures of executive functions did not load on the same factor and the updating function loaded on the same factor with operation span. The researcher then decided to combine operation span and updating and treat the other two executive functions separately. A relatively new approach is to examine the indirect effects of working memory. Using structural equation modeling, Kim et al. (2021) found that working memory predicted literacy, which in turn predicted writing quality, and that inhibition predicted fluency, which in turn predicted writing quality. Zabihi (2018) used path analysis to examine the relationship between anxiety, self-efficacy, and verbal working memory on one hand and writing outcomes on the other. The study showed that all three predictors had direct effects on outcome measures and that self-efficacy also had an indirect effect on writing via anxiety. These two studies exemplify an analytic approach to examining the direct and indirect effects of working memory. This approach also fits the mechanism through which working memory may influence L2 writing on the grounds that it may have a direct effect on writing performance during the writing task and an indirect effect on writing outcomes by contributing to long-term memory such as L2 knowledge, content knowledge, genre knowledge, etc. Furthermore, it must be clarified that results based on path analysis and structural equation modeling cannot be used to make claims about causal relationships, even though a change in the predictor variable may lead to a change in the outcome variable (Collier, 2020). To claim a causal relationship, one would need to use an experimental design. Therefore, whether a causal relationship can be claimed depends on the research design rather than the statistical analysis.

Transparent reporting has a direct effect on the replicability, and credibility of empirical research is critical to scientific research. In this data set, some critical statistical indexes were missed in some studies, such as reliability and statistical assumptions. No study reported performing a power analysis to determine the sample size before a study was conducted or the likelihood of finding significant results based on the sample size of the study after it was conducted. For example, Kormos and Sáfár (2008) did not find a significant correlation between verbal working memory and L2 writing, but the power was only .36, which means that the likelihood of finding a significant effect based on the current sample ( $N = 45$ ) and obtained result ( $r = .19$ ) is only 36%. Transparent reporting is not restricted to statistical analysis, and it relates to all aspects of the methodology of an empirical study including instruments, materials, coding, scoring, and procedure, which were underreported in the primary research, as discussed in previous sections.

### *Findings on the role of working memory*

The predictive power of verbal working memory on overall L2 writing performance is inconsistent and largely nonexistent. This finding can be interpreted in several ways. First, considerable sampling heterogeneity is evident in the studies in this data set and is likely a cause for failure to obtain hypothesized results. The ESL learners in Michel et al. (2019) were from two grade levels and two elementary schools, the EFL learners in Zabihi (2018) were from three universities, and the English language learners in Peng et al. (2022) were from three grade levels and three large school districts. Second, global measures of writing are likely not sensitive enough to detect working memory effects. This speculation is indirectly supported by the fact that more specific measures such as CAF measures were found to be significantly related to working memory in some of the same studies that used global measures of writing such as Révész et al. (2017) and

Vasylets and Marín (2020). As Vasylets and Marín (2020) observed, CAF measures “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” (p. 10). Zalbidea (2017) made a similar comment: “Most studies have failed to supplement global indices of complexity and accuracy with task-specific linguistic measures” (p. 336). Also, as discussed in previous sections, the assessment of overall performance has been conducted in myriad ways and often lacks theoretical basis and empirical evidence for test validity, which may have contributed to the lack of significant results. Notwithstanding the overall lack of significant findings, two studies did find significant correlations between working memory and overall writing performance (Leong et al., 2019; Peng et al., 2022). Leong et al. (2019) adopted a more nuanced approach investigating the moderating effect of genre, and Peng et al. (2022) used a latent-factor approach where working memory and the outcome variable were represented by multiple indicators. The implication is that significant effects of working memory on overall writing may become evident if more rigorous methods are used.

It is also possible that verbal working memory is indeed not involved in overall writing quality, especially in the absence of cognitive burden during the writing process, as happens in untimed writing (Manchón et al., 2023, in this issue) or writing tasks L2 learners typically complete as homework assignments. However, it can be argued that offline writing tasks may still involve working memory resources and that the role of working memory may be indirect via other aspects such as linguistic knowledge, which working memory may have contributed to prior to the writing task. One anonymous reviewer suggested distinguishing writing development and writing performance, with the former referring to learners’ improvement in their writing skills between two points and the latter to task performance at a fixed point. The argument is that working memory is more likely involved in writing development than in writing performance, which is subject to a multitude of factors, thereby eclipsing the role of working memory. Clearly, whether working memory has differential associations with writing development and writing performance is an empirical question.

Verbal working memory has been found to be predictive of CAF measures. The results will be further discussed in the section below on factors moderating the effects of working memory, but several findings are highlighted here. First, verbal working memory may have a negative role in writing accuracy, as shown by Zabihi (2018), which might be attributable to the time pressure of the writing task—writers had no planning time and were required to complete within 11 min. It would seem that in pressured writing tasks, writers with greater verbal working memory prioritize complexity while sacrificing accuracy and those with lesser working memory resources prioritize accuracy over complexity. Second, the findings of Révész et al. (2017) and Vallejos (2020) on fluency are inspiring: Stronger verbal working memory led to less frequent interparagraph pauses but more frequent between-sentence pauses. It would be interesting to explore what these results mean, especially in combination with the finding on inhibition, which is to be discussed below—namely, that writers who are better at inhibiting irrelevant information paused for shorter periods (Kim et al., 2021). Pausing is sometimes considered a negative indicator of writing quality/ability, but in light of these findings, perhaps it is necessary to distinguish pause frequency and pause length. It would seem that more working memory resources enable writers to pause more frequently but for a shorter duration. Note further that pausing is a primary construct in written composition because it has been investigated as both a process and product aspect of writing, it has been used to investigate all subprocesses of writing, and

it has been found to be a consistent predictor of L2 writing quality (e.g., Kim et al., 2021).

Compared with verbal working memory, other components or types of working memory are much less researched. The studies on phonological and visual-spatial working memory are discussed below. The limited research on executive functions showed that inhibition may have a positive effect on minimizing L1 transfer (Arfé & Danzak, 2020) and enable writers to pause for a shorter period (Kim et al., 2021). Updating was examined by Kim et al., who found no significant results, which is perhaps partly attributable to the participants' heterogeneous backgrounds: They were ESL learners from different majors who had stayed in the United States for varied numbers of years; some were international students and others were not.

In the following section, I discuss some moderating factors for the effects of working memory based on the synthesized studies and other related studies. These are preliminary findings or suggestive patterns that may inspire further research; they are by no means conclusive.

### *Proficiency*

One emergent theme is the moderating effects of learners' proficiency level on the role of working memory in writing. Kormos and Sáfár (2008) found a significant effect for phonological short-term memory at an intermediate level but not a beginning level. They speculated that phonological short-term memory is likely more important for implicit learning that occurs at advanced stages of learning, whereas learning at beginning stages is likely more explicit. Vasylets and Marín (2020) showed that verbal working memory was predictive of low-proficiency learners' accuracy and high-proficiency learners' lexical sophistication. Gunnarsson-Largy et al. (2019) found that lower level learners draw primarily on visuospatial working memory, whereas higher level learners draw on both phonological short-term memory and visual-spatial working memory during translation. The findings of these studies are inspiring, and they may serve as hypotheses for further research on the moderating effects of proficiency on the role of working memory in L2 writing.

### *Planning type*

Planning concerns two aspects of composition: content and organization. Skilled writers plan organization more frequently than content, and less skilled writers do the opposite (Révész et al., 2017; Vallejos, 2020). Therefore, it is possible that working memory has a positive correlation with organization planning and a negative correlation with content planning. Note that in writing models, planning only concerns the content and organization of composition and it is separate from translation or formulation. This is different from how planning is conceptualized or examined in L2 speech research, where planning entails both content and language (Li & Fu, 2018).

### *Target structure*

Zalbidea (2017) found that verbal working memory predicted the accuracy of gender agreement but not number agreement errors. The researcher speculated that this is likely because English has number agreement but not gender agreement, which may have posed a greater challenge than number agreement. Note further that gender agreement poses other challenges for L2 learners such as lack of semantic value, multiple form-meaning relations (opacity), and salience. The study suggests the

fruitfulness of taking a fine-grained approach to investigating the role of working memory in learners' use of specific linguistic structures. Of course, there need to be theoretical grounds for the investigation of the moderating effects of linguistic structures, which can be classified based on salience, complexity, developmental sequence, etc.

### *Task demands*

Task demands can be divided into two types: cognitive and procedural demands. Cognitive demands relate to the information processing dimensions of a writing task such as making inferences, providing facts, explaining the process, telling a narrative, etc. The two studies on cognitive demands operationalized as task complexity showed different findings: Cho (2018) failed to find any effects for verbal working memory, whereas Zalbidea (2018) reported a significant correlation between verbal working memory and accuracy in a complex task condition but not in a simple task condition. Zalbidea's study seems to have confirmed Robinson's (2011) prediction that individual difference factors are more likely to be implicated in complex tasks than in simple tasks. Zalbidiea's study validated task complexity, and the writing tasks in the different conditions were strictly controlled, which may have made it more likely for significant results to emerge. Joining scholars in L2 task-based research (Baralt, 2013; Sasayama, 2016), I make a call for validating task complexity and other independent variables by collecting independent evidence—evidence that is not related to the outcome variable or L2 performance.

Genre differences can be approached from the perspective of cognitive demands in that different genres may pose different kinds and amounts of cognitive demands. Genre has been found to be a moderator of working memory effects in L2 writing. Leong et al. (2019) showed that working memory was a significant predictor of explanation (expository writing), a near-significant predictor of argumentation, but a nonsignificant predictor of narration. Therefore, working memory and its components may have differential effects of writing in different genres. In L1 writing, visual-spatial working memory was found to be predictive of descriptive writing but not argumentative writing (Olive, 2022). These findings are suggestive of the importance of examining the interaction between type of working memory and genre. However, genre should not be equated with task complexity. Although it is generally assumed that argumentative writing is more cognitively demanding than narrative writing, this is an assumption. It can be argued that narrative writing based on stipulated content such as recreating a narrative based on a video or a set of pictures may pose greater cognitive demands than writing an argumentative essay where the writer has more flexibility in the choice of content as well as linguistic resources. Thus, the amount of cognitive load of a writing task must be evaluated empirically, as emphasized above.

Different from cognitive demands, procedural demands pertain to the way a task is implemented such as whether learners are allowed to plan before writing, whether they have access to reference tools or corpus materials, whether a time limit is imposed, or whether there is a word limit. In task-based research, it was found that planning made a difference in the role of working memory in L2 speech performance. For example, in Li and Fu (2018), L2 Chinese learners were divided into two groups: within-task planning and pretask planning. In the within-task planning condition, speakers had unlimited time for task completion and were encouraged to plan during task performance, whereas in the pretask planning condition learners had 10 min to plan before task performance but they were given a time limit and were pressured to complete the task

within 5 min. The study results revealed that working memory was involved in within-task planning but not pretask planning. To date, there has been no research on the influence of procedural task demands on the role of working memory in L2 writing.

### *Instructional intervention*

Although Hayes' (Hayes & Flower, 1980; Hayes, 1996) model posits two major categories of contributing factors for writing—learner-related and environment-related factors—most psychological research has focused on learner-related factors investigating the cognitive processes of writing and there has been little attention to the influence of environment-related factors (see, for example, Li, 2017b) or the interaction between the two types of factors. In L2 research, the participants are primarily students in language classes where the effects of instruction on the students' improvement of writing and language ability are of primary interest to practitioners. Therefore, the focus of the research needs to be adjusted. To date, Li and Roshan (2019) is the only study examining how working memory fares in different types of written feedback—an instructional device that has received much attention in L2 research. They found that verbal working memory was positively predictive of the effectiveness of metalinguistic feedback but phonological short-term memory was a negative predictor of the effectiveness of direct correction plus revision. Examining the interaction between types of writing instruction and working memory informs pedagogical decisions and therefore enhances ecological validity.

### *Visual-spatial working memory*

The role of visual-spatial working memory is understated and underresearched. The involvement of visual-spatial memory may happen during planning if the content is concrete (Kellogg et al., 2007); during translation—grammatical encoding, especially syntactic encoding that involves the linear relationship between sentence elements and their positioning, and orthographic encoding; during transcription—programming and typing; and during revision—keeping orthographic and a visual representation of the sentence in mind when making local changes and keeping the locations of related information in mind so the local change is aligned with the big picture and with previous and subsequent parts. There may be an intricate relationship between visual-spatial working memory and phonological short-term memory. For example, referring to previous research, Gunnarsson-Largy et al. (2019) concluded that “visual WM is involved in the conceptualization subprocess, whereas phonological WM is generally involved in the formulation subprocess” (p. 2084). Furthermore, there may be a need to separate visual and spatial working memory, as there is evidence showing their differential roles in the writing process (Kellogg et al., 2013).

## **Conclusion**

This synthetic review examined the theory and research on the role of working memory in second language writing. Overall, there has been limited research; existing research displayed a high degree of methodological heterogeneity; and the findings are equivocal, inconsistent, and at times contradictory. The 16 studies that have been conducted have mostly adopted a regression approach, and there has been one study using the dual-task method. They mainly focused on verbal working memory, and there is insufficient attention to phonological short-term memory, visual-spatial working

memory, and executive functions. The bulk of the research targeted university L2 learners, and other learner populations are underresearched. The samples were heterogeneous, which may have confounded the results and led to unexpected findings. The studies mostly used a product-based approach investigating working memory's associations with overall writing performance and CAF, and there needs to be more process-based research investigating the involvement of working memory in the sub-processes of writing such as planning, translating, transcribing, and editing. The measurement of working memory and writing outcomes needs clarification, validation, and justification before reaching any firm conclusions about the examined research questions. The studies exemplify the fruitfulness in using certain methods of data elicitation for writers' behaviors and cognitive processes, such as keystroke logging, stimulated recall, and eye tracking.

The findings suggest that working memory is largely unrelated to overall writing performance, but the two studies that used more nuanced and sophisticated approaches obtained significant results. Therefore, the lack of importance of working memory in overall writing performance can be concluded only if the rigor of research methods is assumed. Compared with overall proficiency, CAF measures were more likely to show significant links with working memory measures. Despite the small amount of research and heterogeneous methods, the findings are suggestive of theoretically meaningful patterns, promising perspectives, and inspirational directions. More specifically, the role of working memory and L2 writing has been shown to be or is potentially moderated and mediated by learner- and task-related factors (Hayes, 1996) including, but not limited to, proficiency level, genre, the target structure, task demands pertaining to cognitive and procedural dimensions of writing tasks, and instructional interventions. There is also a need to investigate other individual difference factors such as motivation, anxiety, self-efficacy, language aptitude, etc. to unveil the joint and unique contributions of different learner traits and dispositions to L2 writing processes and outcomes. This research would contribute to the general trend in SLA research toward more attention to the role of individual difference factors in L2 learning (Li et al., 2022).

## References

(\*included in the synthesis)

- Adams, A., Simmons, F., Willis, C., & Pawling, R. (2010). Undergraduate students' ability to revise text efficiently: Relationships with topic knowledge and working memory. *Journal of Research in Reading*, 33, 54–76.
- \*Arfé, B., & Danzak, R. L. (2020). The influence of first language spelling and response inhibition skills on English-as-an-additional-language spelling. *Cognitive Development*, 56, 1–14.
- Baddeley, A. (2015). Working memory in second language learning. In Z. Wen, M. Mota, & A. McNeil (Eds.), *Working memory in second language acquisition and processing* (pp. 17–28). Multilingual Matters.
- Baddeley, A. (2017). Modularity, working memory and language acquisition. *Second Language Research*, 33, 299–311.
- Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105, 158–173.
- Baralt, M. (2013). The impact of cognitive complexity on feedback efficacy during online versus face-to-face interactive tasks. *Studies in Second Language Acquisition*, 35, 689–725.
- \*Cho, M. (2018). Task complexity, modality, and working memory in L2 task performance. *System*, 72, 85–98.
- Collier, J. (2020). *Applied structural equation modelling using AMOS*. Routledge.

- Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.
- Engle, R. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11, 19–23.
- Galbraith, D., Ford, S., Walker, G., & Ford, J. (2005). The contribution of different components of working memory to knowledge transformation during writing. *L1: Educational Studies in Language and Literature*, 5, 113–145.
- \*Gunnarsson-Largy, C., Dherbey, N., & Largy, P. (2019). How do L2 learners and L1 writers differ in their reliance on working memory during the formulation subprocess? *Reading and Writing*, 32, 2083–2110.
- 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., & Chenoweth, N. A. (2006). Is working memory involved in the transcribing and editing of texts? *Written Communication*, 23, 135–149.
- Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3–30). Erlbaum.
- 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.
- Kellogg, R. T., Olive, T., & Piolat A. (2007). Verbal, visual and spatial working memory in written language production. *Acta Psychologica*, 124, 382–397.
- Kellogg, R. T., Whiteford, A. P., Turner, C. E., & Cahill, M. (2013). Working memory in written composition: An evaluation of the 1996 model. *Journal of Writing Research*, 5, 159–190.
- Kim, Y., & Schatschneider, C. (2017). Expanding the developmental models of writing: A direct and indirect effects model of developmental writing (DIEW). *Journal of Educational Psychology*, 109, 35–50.
- \*Kim, M., Tian, Y., & Crossley, S. A. (2021). Exploring the relationships among cognitive and linguistic resources, writing processes, and written products in second language writing. *Journal of Second Language Writing*, 53, 1–15.
- Kim, Y. (2022). Do written language bursts mediate the relations of language, cognitive, and transcription skills to writing quality? *Written Communication*, 39, 200–227.
- \*Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory, and foreign language performance in intensive language learning. *Bilingualism: Language and Cognitive*, 11, 261–271.
- Larigauderie, P., Guignouard, C., & Olive, T. (2020). Proofreading by students: Implications of executive and non-executive components of working memory in the detection of phonological, orthographical, and grammatical errors. *Reading and Writing*, 33, 1015–1036.
- Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze writing processes. *Written Communication*, 30, 358–359.
- \*Leong, C. K., Shum, M. S. K., Tai, C. P., Ki, W. W., & Zhang, D. (2019). Differential contribution of psycholinguistic and cognitive skills to written composition in Chinese as a second language. *Reading and Writing*, 32, 439–466.
- Levitt, W. (1989). *Speaking: From intention to articulation*. The MIT Press.
- Levy, C. M., & Marek, P. (1999). Testing components of Kellogg's multicomponent model of working memory in writing: The role of the phonological loop. In M. Torrance & G. C. Jeffery (Eds.), *The cognitive demands of writing: Processing capacity and working memory in text production* (pp. 25–41). Amsterdam University Press.
- Li, S. (2023). Working memory and second language learning: A critical and synthetic review. In Godfroid & H. Hopp (Eds.), *The Routledge handbook of second language acquisition and psycholinguistics* (pp. 348–360). Routledge.
- Li, S. (2017a). Cognitive differences and ISLA. In S. Loewen & M. Sato (Eds.), *Handbook of instructed second language learning* (pp. 396–417). Routledge.
- Li, S. (2017b). The effects of cognitive aptitudes on the process and product of L2 interaction: A synthetic review. In L. Gurzynski (Ed.), *Expanding individual difference research in the interaction approach: Investigating learners, instructors and researchers* (pp. 41–70). John Benjamins.
- Li, S., & Fu, M. (2018). Strategic and unpressured within-task planning and their associations with working memory. *Language Teaching Research*, 22, 230–253.

- Li, S., Hiver, P., & Papi, M. (2022). Individual differences in second language acquisition: Theory, research, and practice. In S. Li, P. Hiver, & M. Papi (Eds.), *The Routledge handbook of SLA and individual differences* (pp. 3–34). Routledge.
- \*Li, S., & Roshan, S. (2019). The associations between working memory and the effects of four different types of written corrective feedback. *Journal of Second Language Writing, 45*, 1–15.
- \*Lu, Y. (2010). *Cognitive factors contributing to Chinese EFL learners' L2 writing performance in timed essay writing* [Unpublished doctoral dissertation, Georgia State University].
- Manchón, R., McBride, S., Mellado Martínez, M. D., & Vasylets, O. (2023). Working memory, L2 proficiency, and task complexity: Independent and interactive effects on L2 written performance. *Studies in Second Language Acquisition, 45*, 737–764.
- Manchón, R., & Polio, C. (Eds.) (2022). *The Routledge handbook of second language acquisition and writing*. Routledge.
- \*Mavrou, I. (2020). Working memory, executive functions, and emotional intelligence in second language writing. *Journal of Second Language Writing, 50*, 1–13.
- \*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.
- Miyake, A., & Friedman, N. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*, 8–14.
- Olive, T. (2022). How do novice and skilled writers engage working memory? In J. Schwieter & Z. Wen (Eds.), *Working memory and language* (pp. 504–525). Cambridge University Press.
- \*Peng, A., Orosco, M. J., Wang, H., Swanson, H. L., & Reed, D. K. (2022). Cognition and writing development in early adolescent English learners. *Journal of Educational Psychology, 114*, 1136–1155.
- \*Révész, A., Michel, M., & Lee, M. (2017). *Investigating IELTS academic writing task 2: Relationship between cognitive writing processes, text quality, and working memory*. IELTS Research Reports Online Series. [https://www.ielts.org/en-us/for-researchers/research-reports/ielts\\_online\\_rr\\_2017-3](https://www.ielts.org/en-us/for-researchers/research-reports/ielts_online_rr_2017-3)
- 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*. John Benjamins.
- Sagarra, N. (2017). Longitudinal effects of working memory on L2 grammar and reading abilities. *Second Language Research, 3*, 341–363.
- Salas, N., & Silvente, S. (2020). The role of executive functions and transcription skills in writing: A cross-sectional study across 7 years of schooling. *Reading and Writing, 33*, 877–905.
- Sasayama, S. (2016). 'Is "complex" task really complex? Validating the assumption of cognitive task complexity.' *Modern Language Journal, 100*, 231–254.
- Shah, P., & Miyake, P. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology, 125*, 4–27.
- Swanson, H., & Berninger, V. (1996). Individual differences in children's writing: A function of working memory or reading or both processes? *Reading and Writing, 8*, 357–383.
- \*Vallejos, C. A. (2020). *Fluency, working memory and second language proficiency in multicompetent writers* [Doctoral dissertation, Georgetown University]. Georgetown University Library Repository. <https://repository.library.georgetown.edu/handle/10822/1059667>
- Vandenberg, R., & Swanson, H. L. (2007). Which components of working memory are important in the writing process? *Reading and Writing, 20*, 721–752.
- Vasylets, O., & Marin, J. (2020). The effects of working memory and L2 proficiency on L2 writing. *Journal of Second Language Writing, 52*, 1–14.
- Wen, Z., & Li, S. (2019). Working memory in L2 learning and processing. In J. Schwieter & A. Benati (Eds.), *The Cambridge handbook of language learning* (pp. 365–389). Cambridge University Press.
- \*Zabihi, R. (2018). The role of cognitive and affective factors in measures of L2 writing. *Written Communication, 35*, 32–57.
- \*Zalbidea, J. (2017). One task fits all? The roles of task complexity, modality, and working memory capacity in L2 performance. *The Modern Language Journal, 101*, 335–352.

---

**Cite this article:** Li, S. (2023). Working memory and second language writing: A systematic review. *Studies in Second Language Acquisition, 45*: 647–679. <https://doi.org/10.1017/S0272263123000189>