https://doi.org/10.69760/jales.2026001003

Visuospatial Working Memory Load Effects on Predictive Eye Gaze in L2 Phonological Processing

[1] Zarifa Sadigzade

Abstract

Second language (L2) processing research has predominantly focused on verbal working memory (WM), yet the potential role of visuospatial WM (VSWM) remains underexplored. This study examines how a concurrent VSWM load impacts predictive eye movements during L2 spoken word processing. Fifty adult L2 English learners completed a visual-world eye-tracking experiment in which they listened to sentences that were either predictive or non-predictive of an upcoming noun, while simultaneously performing a secondary visuospatial memory task (symmetry span) in a high-load condition. Growth curve analyses revealed that under high VSWM load, anticipatory fixations to target objects were delayed by approximately 150 ms compared to a low-load baseline, a significant effect ($\beta = -0.22$, $p < 0.001$). No interaction with L2 proficiency was observed. These findings suggest that VSWM capacity constrains real-time phonological prediction in L2 listening, extending theoretical models of WM in SLA and offering practical insights for multimedia language learning.

Keywords

Visuospatial working memory, Predictive processing in L2 listening, Eye-tracking

Introduction

Working memory (WM) is widely recognized as a key individual difference in second language (L2) acquisition and processing. Classic models of L2 proficiency have emphasized verbal WM components—particularly the phonological loop and executive control—as primary drivers of L2 learning success (Wen, 2012; Linck, Osthus, Koeth, & Bunting, 2014). For example, Wen (2012) outlined a "phonological/executive" model of WM in L2 learning, positing that phonological short-term memory and executive attention jointly underpin learners' capacity to acquire and process an L2. In support of this emphasis, Linck et al. (2014) conducted a comprehensive meta-analysis and found a reliable positive correlation (population $\rho \approx 0.25$) between WM capacity and L2 processing skill. Intriguingly, this WM–L2 link held across both lower- and higher-proficiency learners, suggesting that even advanced L2 users' processing efficiency is influenced by their WM resources. Notably, complex span (executive) measures showed stronger associations with L2

[1] Zarifa Sadiqzade, lecturer, Nakhchivan State University, zarifasadig@gmail.com, ORCID: https://orcid.org/0009-0007-1179-1214

This is an open access article under the **Creative Commons Attribution 4.0 International License**

outcomes than simple verbal memory measures, indicating the dominance of the verbal executive aspect of WM in prevailing L2 processing models. In sum, prior research solidly implicates verbal WM—both phonological short-term storage and executive attention—in myriad L2 skills ranging from vocabulary learning to reading and speaking.

However, this traditional focus on verbal WM overlooks the fact that WM is a multi-component system including a visuospatial sketchpad (Baddeley & Hitch, 1974; Baddeley, 2000). According to Baddeley's multicomponent model, WM comprises at least four components: a phonological loop for sound-based information, a visuospatial sketchpad for visual/spatial data, a central executive for attentional control, and an episodic buffer for multimodal integration. While the phonological loop and central executive have been studied extensively in relation to language, the visuospatial WM component has received scant attention in L2 research. This gap is striking because cognitive studies suggest that visuo-spatial memory capacity is a distinct construct that can significantly affect task performance independent of verbal memory. For instance, Cornoldi and Vecchi (2003) proposed an expanded WM framework with two dichotomies (verbal vs. visuospatial, and passive storage vs. active processing). In their view, retaining visual information (e.g. shapes or locations) engages memory resources separate from those used for verbal material, especially when active manipulation is required. Empirical work supports this separation: Giofrè et al. (2013) found that adolescents' performance in geometry was uniquely predicted by their active visuospatial WM capacity, whereas simple visual storage was less predictive. Moreover, children with nonverbal learning disabilities (i.e. visuo-spatial deficits) show marked difficulties in intuitive geometry, and these group differences are largely explained by poorer VSWM (complex span) abilities. Such findings underscore that visuospatial memory is not merely a peripheral cognitive resource, but rather a capacity that, when taxed, can constrain complex cognitive operations. By extension, it is plausible that VSWM might also play a role in language tasks that involve visual contexts or cues – a hypothesis that remains virtually untested in SLA to date.

Motivated by these theoretical and empirical gaps, the present study targets the underexplored intersection of VSWM and L2 processing. Specifically, we ask: Does loading the visuospatial sketchpad interfere with L2 learners' ability to anticipate upcoming linguistic information during listening? Real-time language comprehension often involves prediction, the proactive pre-processing of likely upcoming words or structures. In native (L1) processing, ample evidence shows that listeners and readers can use context to predict specific upcoming words (e.g., Altmann & Kamide, 1999). In L2 processing, however, the picture is mixed. Some accounts suggest that L2 learners engage in less prediction or at least different predictive strategies compared to native speakers (Kaan, 2014; Kaan, Kirkham, & Wijnen, 2014). For example, Kaan (2014) reported that advanced L2 readers did not show the same pre-activation of upcoming syntactic information (ellipsis sites) that natives did, even though both groups were sensitive to the context after the fact.

This is an open access article under the Creative Commons Attribution 4.0 International License

This implies that L2 speakers might rely more on integration "after the event" and less on proactive prediction, possibly due to processing constraints. The exact nature of those constraints is debated: Are L2 users limited by reduced processing capacity (Just & Carpenter, 1992), or by greater interference and less efficient cue use in memory retrieval (Cunnings, 2017)? Cunnings (2017) has argued that L2 comprehenders may be especially susceptible to similarity-based interference in memory, meaning they might retrieve wrong antecedents or lexical items because multiple candidates compete, partly owing to how L2 learners weight cues. He suggests L2 learners rely more on discourse cues, which could lead to interference when those cues mislead retrieval. In contrast, capacity-based accounts maintain that limitations in attentional resources or storage capacity can directly hinder simultaneous processing and anticipation (Just & Carpenter, 1992). Under this view, L2 learners might predict less because managing multiple information sources (e.g. processing context, maintaining the input, generating predictions) exceeds their available WM resources, especially for less proficient learners.

Critically, both perspectives highlight working memory as a central factor, but they differ in mechanism. To date, most evidence on WM and L2 sentence processing is correlational or based on offline measurements (e.g., relating complex span scores to global comprehension success). Few studies have directly manipulated cognitive load during online L2 processing to examine causality. An exception is Ito, Corley, and Pickering (2018), who used a dual-task paradigm to investigate the effect of a secondary verbal memory load on predictive eye movements. They found that both L1 and L2 speakers showed delayed anticipatory eye gaze when performing a concurrent word memorization task. In Ito et al.'s study, participants listened to sentences with strongly constraining verbs (e.g., "The boy will eat the…") while viewing a display of objects; normally, listeners would begin looking at the likely target (e.g., a cake) before it is named. When participants had no extra task, this predictive looking emerged early; with an added memory load, the rise in target fixations was significantly delayed (by roughly 200 ms). Importantly, Ito et al. reported that the pattern was similar for L1 and L2 listeners, suggesting that L2 speakers, when matched in proficiency, utilize prediction in fundamentally the same way as natives, provided they have sufficient cognitive resources. Their findings support the idea that making predictions is resource-dependent.

While Ito et al.'s study shed light on the role of verbal WM load, it also raises new questions. Their secondary task tapped verbal storage (memorizing words), leaving open whether a load on visuospatial resources would have an analogous impact. On one hand, if predictive processing in language primarily draws on domain-general executive attention, any significant concurrent task – even in a different modality – could siphon resources away and thus delay or reduce prediction. On the other hand, if prediction is more specifically tied to phonological rehearsal or language-specific memory, a visuospatial task might interfere less (or differently) than a verbal task. Furthermore, almost no prior studies have examined phonologically-driven predictive eye

This is an open access article under the Creative Commons Attribution 4.0 International License

movements in L2. Most L2 visual-world eye-tracking research has focused on semantic or syntactic cues (e.g., verb semantics, case markings, gender agreement) rather than purely phonological anticipation. This is a notable gap: in natural listening, listeners continuously map unfolding sounds onto potential words, and might anticipate upcoming words based on initial phonemes and contextual constraints. Whether L2 learners can do so efficiently – and how WM limitations might affect the time-course of phonological processing – remains under-investigated. We also note the lack of research integrating WM load with phonological prediction specifically; thus, our study extends previous dual-task designs by focusing on phonological input processing under load.

In summary, there is a confluence of gaps in the literature: (1) Visuospatial WM's role in L2 processing is underexplored, (2) the effect of cognitive load on predictive processing in L2 (especially at the phonological level) is not well understood, and (3) more generally, the field has called for approaches that can disentangle whether WM effects in L2 are due to capacity limitations or interference (Cunnings, 2017). The present study addresses these gaps by employing a dual-task paradigm to explicitly impose a visuospatial WM load during an L2 predictive processing task. By observing changes in L2 learners' anticipatory eye gaze under high vs. low VSWM load, we can infer the extent to which domain-general resources versus domain-specific interference constrain L2 prediction. We also examine whether L2 proficiency modulates the load effect, given prior suggestions that higher proficiency might mitigate WM constraints (or conversely, that WM impacts persist even at advanced levels). Through this investigation, we aim to contribute novel evidence on how the often-neglected visuospatial component of WM figures into language processing, thereby broadening the theoretical understanding of WM in SLA (Wen, 2012) and informing practical considerations for multimodal language learning scenarios.

Research on working memory in second language acquisition has traditionally centered on verbal WM systems. These include phonological short-term memory (PSTM) – the ability to temporarily store and rehearse sound-based information – and the central executive or executive WM, responsible for attentional control and manipulation of information. Numerous studies link these verbal WM components to L2 learning outcomes. For instance, Alptekin and Erçetin (2010) examined the role of L1 vs. L2 WM in L2 reading comprehension. They found that both a reader's L1 WM span and L2 WM span contributed to understanding L2 texts, particularly for inferential comprehension questions that required integrating information across sentences. This suggests that a good memory for language (in either L1 or L2) aids higher-level text processing. Notably, Alptekin and Erçetin observed that differences between WM measured in L1 and L2 diminished as proficiency increased. In other words, more proficient L2 readers approached a point where their memory capacity in L2 tasks was comparable to that in L1, aligning with the idea that increased automaticity in L2 frees up WM resources (or that advanced learners can effectively

This is an open access article under the Creative Commons Attribution 4.0 International License

harness L1 cognitive resources for L2 tasks). These findings reinforced the significance of verbal WM in L2 reading and hinted at complex interactions with proficiency.

Beyond reading, verbal WM has been implicated in L2 grammar learning, vocabulary acquisition, and oral production. Wen (2012) highlighted that phonological memory (often measured by nonword repetition or digit span) is a strong predictor of L2 vocabulary acquisition and early-stage grammar learning, as it helps learners retain novel phonological sequences. The phonological loop is seen as a "language learning device" (Baddeley, Gathercole, & Papagno, 1998) that is crucial for encoding new words. On the other hand, Linck et al. (2014) synthesized 79 studies and confirmed a modest but consistent correlation between WM capacity and a range of L2 processing and proficiency measures ($r \sim .25$). Importantly, their analysis indicated that complex span tasks, which tax both storage and processing (e.g., reading span, operation span), were better predictors of L2 performance than simple span tasks. This aligns with findings in L1 reading research (Daneman & Merikle, 1996) and suggests that the executive control aspect of WM – the ability to maintain information while engaging in concurrent processing – is particularly relevant for complex L2 tasks like understanding sentences and discourse. In fact, Linck et al. found that the WM–L2 link was not restricted to beginners; even highly proficient L2 users showed performance differences attributable to WM capacity. This challenges any assumption that advanced learners "transcend" cognitive limitations, and echoes other work (e.g., Serafini & Sanz, 2016; Cunnings, 2017) suggesting that individual differences can persist across proficiency levels.

The preoccupation with verbal WM in SLA is also reflected in theoretical models. Wen's (2015, 2016) phonological/executive model explicitly foregrounds two WM components: a phonological buffer (for sound-based storage) and an executive control mechanism. According to Wen (2016), these two components are most germane to L2 learning and use. Phonological WM supports the acquisition of new words and formulas (especially in early stages), whereas executive WM becomes crucial for complex tasks like reading comprehension, conversation, and using feedback, particularly at higher proficiencies. This dichotomy dovetails with empirical findings: early in L2 acquisition, simple memory span (PSTM) correlates with vocabulary uptake, while in later stages, tasks requiring attention-switching and inhibition (as indexed by complex spans) better predict performance (Wen & Li, 2019). Overall, the literature firmly establishes verbal WM as a cornerstone of L2 aptitude and processing. Yet, this focus has also led to a notable gap – the relative neglect of the visuospatial aspect of WM, which we address next.

Visuospatial Working Memory: The Missing Piece

Visuospatial working memory (VSWM) refers to the ability to temporarily hold and manipulate visual and spatial information. In Baddeley's model, this is the "sketchpad" that parallels the phonological loop. It handles information such as locations, shapes, and visual patterns. Mainstream SLA research has seldom considered VSWM, perhaps assuming that language

This is an open access article under the Creative Commons Attribution 4.0 International License

processing is predominantly verbal. However, as language use is often situated in visual contexts (e.g., conversations happen in environments, listening frequently coincides with looking at scenes or faces, reading involves visual text), there are credible pathways for VSWM to influence L2 processing. Furthermore, some learners may rely on visualization strategies or mental imagery when processing or learning language, linking visual memory with verbal tasks (e.g., remembering a word by picturing its referent).

Support for the potential importance of VSWM comes from cognitive psychology and educational research. Cornoldi and Vecchi (2003) argued that WM should be conceived along two independent axes: modalities (verbal vs. visuo-spatial) and processing demands (passive maintenance vs. active manipulation). Their model was motivated by findings that some individuals could have high memory spans for one modality but not the other, and that active processing tasks predict outcomes (like problem-solving) better than simple retention tasks. Within the visuo-spatial domain, Cornoldi and colleagues (Cornoldi, Carretti, & De Beni, 2001) found dissociations between "visual" memory (memory for imagery or colors) and "spatial" memory (memory for locations or sequences of moves), further suggesting subcomponents in the sketchpad. The key takeaway is that VSWM is a multifaceted construct with its own capacity limits and sub-processes, which can be specifically taxed by visuo-spatial information.

Empirical evidence underscores VSWM's functional significance. Giofrè et al. (2013) conducted a study with secondary school students solving geometry problems – a task with heavy spatial reasoning demands. They measured various aspects of the students' VSWM using both simple storage tasks (retaining visual patterns or spatial sequences) and complex span tasks (retaining such information while performing another activity). Giofrè et al. found that students' active VSWM capacity (performance on complex spans) was a strong predictor of geometry achievement, whereas simple storage capacity was less predictive. In path analyses, VSWM emerged as a significant contributor to both intuitive geometry ability and formal geometry grades. This implies that when a task (like geometry) requires mentally manipulating spatial information, those with greater VSWM resources have a clear advantage.

Additional evidence comes from neuropsychological studies of learning disabilities. Mammarella et al. (2013) examined children with nonverbal learning disability (NLD), a condition characterized by deficits in visuo-spatial skills despite intact verbal IQ. They discovered that NLD children performed worse than controls on an intuitive geometry task, especially on items involving spatial relations (Euclidean concepts and transformations). Crucially, the NLD group's VSWM performance was significantly lower on complex-span tasks, and these VSWM deficits statistically accounted for their poorer geometry scores. A discriminant analysis confirmed that complex VSWM measures were the best at distinguishing NLD children from typically developing ones. Thus, complex visuo-spatial memory abilities (the kind that involve dynamic processing and

This is an open access article under the Creative Commons Attribution 4.0 International License

maintenance of images in mind) were "crucial" to success in visuo-spatial problem solving. By analogy, one could hypothesize that similar VSWM abilities might aid complex L2 tasks that involve visual contexts—such as using co-speech gestures, interpreting visual scenes in listening tasks, or reading while listening (as in captioned video).

In the L2 domain, direct investigations of VSWM are scarce. One related line of work involves foreign language learning aptitude tests like MLAT, which include components such as memory for spoken syllables or associations of words with pictures. Some aptitude models (e.g., Polychroni et al., 2017) have considered a visuo-spatial memory element in predicting language learning, but empirical results are limited. Another relevant area is research on sign language learning or bimodal bilingualism, where spatial memory might plausibly play a larger role (e.g., remembering sign spatial locations or configurations); however, our focus here is on spoken language in a visual context. The lack of L2 studies explicitly manipulating or measuring VSWM means that we largely infer its importance indirectly. Our study thus breaks new ground by integrating a VSWM task into an L2 processing experiment. If taxing the sketchpad interferes with L2 comprehension, that would provide concrete evidence that visuo-spatial resources are engaged during language processing – a finding that could spur a reevaluation of WM models in SLA to include the whole multicomponent system, not just the phonological loop and central executive.

Eye-Tracking and L2 Predictive Processing

Eye-tracking has become a vital method in investigating L2 processing in recent years (Roberts & Siyanova-Chanturia, 2013; Godfroid, 2019). By recording where and for how long learners direct their gaze, eye-tracking provides a moment-by-moment indication of cognitive processing during reading or listening tasks. Unlike offline measures (e.g., end-of-sentence judgments or comprehension questions), eye movements reflect immediate processing decisions and can capture subtle differences between native and non-native processing. Roberts and Siyanova-Chanturia (2013) stress that eye-tracking allows researchers to study L2 learners' "moment-by-moment interpretation" without interrupting the comprehension process. This is crucial because L2 learners might perform similarly to natives on some end-of-task measures but differ in how they arrived at that understanding. Eye-movement data can reveal these differences, such as prolonged fixations on syntactically complex regions in L2 reading (indicating processing difficulty) or different patterns of regressions (re-reading) when parsing garden-path sentences.

Two main eye-tracking paradigms are used in L2 research: eye-tracking during reading and the visual-world paradigm (eye-tracking during listening with a visual display). The reading paradigm records where on a written text the eyes fixate, yielding measures like first fixation duration, gaze duration, and total time on regions of interest. Using this paradigm, researchers have examined L2 phenomena such as syntactic ambiguity resolution (e.g., Dussias, 2010), anaphora and pronoun processing (Felser & Cunnings, 2012), and lexical frequency effects in L2 vs. L1 reading. For

This is an open access article under the Creative Commons Attribution 4.0 International License

example, in processing filler-gap dependencies (like wh-questions or relative clauses), L2 readers often show longer reading times at the gap position, suggesting greater difficulty integrating the filler with its subcategorizer, and these effects sometimes correlate with WM scores (Juffs & Harrington, 2011). Eye-tracking has also been used to investigate how L2 readers handle misleading cues: Keating (2009) found that less proficient L2 Spanish readers were slower to recover from garden-path sentences, as evidenced by longer go-past times, implying that reanalysis is more costly for them, possibly due to WM limitations.

The visual-world paradigm, on the other hand, is especially suited to studying predictive processing. In this paradigm, participants listen to spoken language while viewing a scene (often with multiple objects). Their gaze tends to shift to referents as they are mentioned, but crucially, it can also shift in anticipation of something being mentioned if the context is constraining. For instance, the seminal study by Altmann and Kamide (1999) showed that when English listeners heard "The boy will eat the…", they started fixating a picture of a cake before the word "cake" was spoken, thanks to the verb "eat" predicting an edible object. Such anticipatory eye movements provide a concrete measure of linguistic prediction in real time. In L2 research, visual-world studies have examined whether L2 learners make similar predictions and under what conditions. Chambers and Cooke (2009) demonstrated that high-context sentences constrained L2 listeners' lexical expectations, reducing interference from unrelated native-language words. They found that when context clearly indicated an upcoming object, L2 listeners (with sufficient proficiency) limited their consideration of competitors, suggesting some degree of prediction or at least rapid integration. However, other studies have found that L2 learners' anticipatory eye movements are weaker or delayed relative to natives. Martin et al. (2013) observed that low-intermediate L2 learners of German did not show prediction based on case-marking cues that reliably signaled the upcoming noun's gender, whereas advanced learners did—but still later than native speakers. These results imply that predictive processing might be a skill that grows with proficiency and exposure, and one sensitive to processing efficiency and WM.

Ito et al. (2018), introduced earlier, is a pivotal study in this area because it directly examined the effect of an extraneous cognitive load on prediction in both L1 and L2. They found that performing a secondary task (memorizing words) slowed down predictive looks in both groups. Interestingly, the magnitude of the delay was similar for L1 and L2 participants, on the order of a few hundred milliseconds. This suggests that given comparable proficiency, L2 listeners can and do predict like natives, but both require spare cognitive capacity to do so. When resources are tied up, the timing of prediction suffers. Kaan (2014) in a review of prediction in L2 noted multiple sources of variability: L2 predictions may be less robust due to factors like weaker linguistic cues, slower processing, or cautious strategies; but also, any observed differences could stem from general processing pressures that L2 users face (like having to devote more effort to lexical retrieval or parsing, leaving less bandwidth for prediction). Kaan emphasizes that not all studies find an

This is an open access article under the Creative Commons Attribution 4.0 International License

absence of L2 prediction—many advanced L2ers do anticipate upcoming information, especially when the cues are transparent and processing is not overloaded. Therefore, a key research direction is to understand under what circumstances L2 learners engage in predictive processing and when those processes break down.

One clear factor is cognitive load. If an L2 listener is juggling too many tasks at once, prediction may be one of the first processes to be curtailed. This is where WM comes into play: individuals with higher WM might handle more information in parallel and maintain predictions even under load, whereas those with lower WM (or when anyone's WM is heavily taxed) might resort to a more reactive mode of processing. The present study leverages the visual-world paradigm to examine this dynamic: by introducing a concurrent visuospatial memory task, we simulate a condition of increased cognitive load and observe how it affects anticipatory eye gaze in L2. This approach is novel in combining an eye-tracking measure of prediction with an online dual-task WM manipulation.

Identified Research Gaps and the Current Study's Contributions

Gap in Literature	Evidence/Source	Addressed by Present Study
Visuospatial WM largely neglected in SLA research. Most L2 studies focus on phonological loop and executive WM, with minimal exploration of the visuospatial sketchpad.	– *Wen (2016)* notes phonological and executive components are considered most relevant, implying other components (VSWM) are overlooked. – *Cornoldi & Vecchi (2003)* propose a WM model including visuo-spatial resources, but SLA has not integrated this.	We incorporate a VSWM load (via a symmetry span task) into an L2 processing experiment. By testing the effect of sketchpad load on comprehension, we directly probe VSWM's role in L2, providing novel data on whether visual memory resources constrain language processing.
Lack of studies on phonological predictive eye gaze in L2. Prior L2 eye-tracking research emphasizes semantic/syntactic cues; it remains unclear if and how L2 learners anticipate upcoming words based on phonological input.	– *Kaan (2014)* reports L2 speakers do not predict upcoming info as much as L1 speakers in some contexts[15] (e.g., less anticipatory ERP effects), but focuses on syntax. – Few visual-world studies test form-based prediction; most use semantic constraints (Ito et al., 2018 used verb semantics). No specific "phonological prediction" studies were found in L2.	Our visual-world design involves phonological processing of unfolding words. We measure how L2 listeners map spoken word onsets to visual referents in real time. By observing anticipatory looks (or their delay) to target objects as the word's initial sounds are heard, we shed light on phonologically-driven prediction in L2.
Limited experimental evidence on WM load in L2 online processing (capacity vs. interference). Debate exists on whether WM effects reflect capacity limitations or interference susceptibility, with few studies manipulating load to test causality in L2.	– *Cunnings (2017)* calls for research to disentangle capacity-based and interference-based accounts of L2 parsing differences. L2 learners may weight cues differently, leading to interference, but this is usually inferred post-hoc. – *Ito et al. (2018)* showed a verbal WM load delays prediction in L2, suggesting capacity effects; no studies with non-verbal load yet.	We implement a dual-task paradigm to causally examine WM's impact. By using a non-verbal (visuospatial) load, we test if general capacity limits (not just linguistic interference) hinder L2 prediction. A significant effect of a spatial memory task on language processing would support capacity-based models, whereas a null effect might imply modality-specific interference is key.
Uncertainty about proficiency moderating WM effects. Some theories suggest high proficiency L2 learners might overcome WM constraints, but meta-analytic evidence is mixed.	– *Linck et al. (2014)* found WM–L2 correlations persist even in advanced learners, contrary to a "threshold" hypothesis. – *Jeon & Yamashita (2014)* (meta-analysis) reported only moderate overall correlation ($r \approx .42$) between WM span and L2 reading, leaving room for proficiency or task factors to mediate the relationship.	We include learners of varying proficiency and explicitly test the Load × Proficiency interaction. Our finding of no proficiency interaction (no differential effect of load on higher vs. lower proficiency) provides evidence that WM constraints operate even at advanced levels. This underscores that proficient L2 users are not immune to capacity limitations under dual-task conditions.

This is an open access article under the **Creative Commons Attribution 4.0 International License**

Table 1: Key gaps identified in previous research and how the present study contributes to filling those gaps.

To summarize, eye-tracking research in L2 has illuminated that (a) L2 processing is often slower and more fixation-intensive, reflecting greater effort; (b) proficient L2 learners can exhibit native-like real-time behaviors, including prediction, under favorable conditions; and (c) cognitive individual differences, such as WM, likely modulate these behaviors (though experimental evidence is still emerging). The gaps identified in prior work—such as the under-investigation of VSWM, the need for more dual-task studies of L2 processing, and questions about L2 predictive mechanisms—set the stage for our study. Table 1 provides a synopsis of key gaps in the literature and how the current research addresses them.

Through the above lenses, our study marries these threads: investigating a novel aspect of WM (visuospatial) in L2 processing, focusing on real-time predictive eye-movements under load, and interpreting the results in light of capacity vs. interference frameworks and proficiency considerations. The next sections detail our methodology, results, and interpretations in this context.

Method

Participants

The participants were 50 adult L2 English speakers (18–35 years old, 35 female, 15 male) recruited from a university population. All were late bilinguals who learned English as a foreign language in classroom settings from around age 10 onward. We set minimum proficiency criteria to ensure participants could understand the spoken materials: all had at least an upper-intermediate English proficiency (B2 or above on the CEFR scale), confirmed by a placement test and self-reports. Within the sample, proficiency ranged from intermediate to advanced; this allowed us to examine whether proficiency modulated the effects of working memory load. None of the participants had known hearing or vision impairments (normal or corrected-to-normal vision), as both auditory and visual acuity were important for the task. Participants either received course credit or a small honorarium for their involvement, and informed consent was obtained in accordance with institutional ethics guidelines.

Materials and Design

We employed a visual-world paradigm to track participants' eye movements as they listened to English sentences. Each trial consisted of a spoken sentence paired with a visual display of four objects on a computer screen. The sentences were designed such that in half of the trials, the sentence context was predictive of a particular object before it was explicitly mentioned (Predictive condition), while in the other half, it was non-predictive or neutral (Control condition). For

This is an open access article under the Creative Commons Attribution 4.0 International License

example, a predictive trial might involve a sentence like “The pirate hides the treasure…” presented with images of a treasure chest, a ship, a palm tree, and a cat. The verb phrase “hides the…” strongly suggests “treasure” as the next word (given the context of a pirate), so an attentive listener could start anticipating “treasure” as soon as they hear “hides the…”. A control version of that item might be “The pirate sees the treasure…” where the verb “sees” does not specifically predict any one object over another, so no strong anticipatory look toward “treasure” would be expected until the word actually begins.

We created 40 item sets like this, rotated across conditions so that each participant saw each context only once (either predictive or control) to avoid repetition effects. The target object (e.g., treasure) and three distractor objects were arranged in distinct quadrants of the screen; distractors were chosen such that one might serve as a phonological competitor in some trials (e.g., another object whose name shares initial sounds with the target) to ensure that anticipatory looks were based on the intended prediction and not random. All images were colored drawings of common objects, roughly equal in visual salience and size, and their names were generally known to learners (concrete nouns with high familiarity). The audio sentences were recorded by a female native English speaker at a natural speaking rate, then normalized for volume. Sentence onset and the timing of critical words (e.g., noun onset) were marked to allow time-locking of eye movement analyses.

Crucially, our design introduced a dual-task manipulation to create a high versus low working memory load condition. The secondary task specifically targeted visuospatial WM using a variant of the symmetry span task (adapted for the visual-world context). In the High Load condition, participants had to remember a sequence of spatial locations while simultaneously processing the sentence. Each high-load trial began with a brief presentation of a 4×4 grid on which one cell was filled (a red square) for 650 ms. Participants were instructed that across the trial they would see several such grids and needed to remember the locations of the filled cells in the exact order they appeared. After the initial grid flash, the visual scene with the four objects would appear, and 500 ms later the sentence audio would play. Following the sentence (and before any response was required about the sentence), the screen would briefly display another filled-cell grid (second in the sequence) for 650 ms, which participants also had to encode. Each trial contained a sequence of two such grid locations to remember (symmetry span length of 2; this length was chosen based on pilot testing to impose load without overwhelming participants to the point of task abandonment). At the end of the trial, after the sentence finished, participants saw an empty 4×4 grid and were prompted to click the cells in the two locations they had seen. This recall step ensured they actively maintained the visuo-spatial information during sentence processing.

In contrast, in the Low Load condition, participants performed no secondary task: they simply listened to the sentence and viewed the objects, with their only objective being to comprehend the

This is an open access article under the Creative Commons Attribution 4.0 International License

sentence. (They were told that occasionally a comprehension question might follow to ensure they paid attention, and indeed we interspersed 10 filler trials with yes/no questions about the sentence content to maintain accountability.)

The experiment thus had a 2 × 2 design: Context Predictability (Predictive vs. Control) × WM Load (High vs. Low). Context was manipulated within subjects (each participant experienced both predictive and control trials for different items), and Load was manipulated between blocks. We used a blocked design for load to minimize frequent task-switching. Half of the participants did the Low Load block first (20 trials) then the High Load block (20 trials), and half did the reverse, to counterbalance any order or practice effects. Within each block, predictive and control sentences were intermixed in a pseudo-random order, with the constraint that no more than two predictive or two control trials appeared consecutively.

Apparatus and Procedure

Eye movements were recorded using an SR Research EyeLink 1000 eye-tracker, sampling at 1000 Hz. Participants were seated approximately 60 cm from a 21-inch monitor. The visual display subtended about 25° of visual angle horizontally; objects were positioned equidistant from the screen center to avoid central bias. We used a 9-point calibration and validation procedure at the start of each block to ensure tracking accuracy within 0.5° visual angle. During trials, gaze position data were sent to the computer and stored for analysis.

At the start of the experiment, participants received instructions and practice. For the symmetry span task, they practiced a few trials of remembering sequences of red-square locations (without any sentences) to familiarize them with the concept. They also completed a couple of practice trials in the dual-task format (seeing grids, hearing a sentence, recalling grids) to ensure they understood the dual-task requirements. In Low Load practice, they were simply told to listen and look at the pictures. The importance of maintaining central fixation until sentence onset was stressed (to avoid anticipatory bias), and they were told that they could freely move their eyes during the sentence "as if watching a scene" – we did not give any strategy for how to look; we simply asked them to listen for comprehension.

During the main task, in High Load trials, participants first saw the grid(s) to remember, then the scene and sentence. After each high-load trial, they performed the recall by clicking the remembered cells on an on-screen grid using the mouse. Feedback on recall accuracy was given to encourage effort (a brief message: "Correct" or "Incorrect, the correct squares were [highlighted]"). We treated the VSWM task primarily as a load manipulation; the recall accuracy data were collected (and later examined to verify that the task was indeed challenging and that participants were engaged), but participants were not excluded for low memory performance as long as they were following instructions. In Low Load trials, a similar temporal structure was

This is an open access article under the Creative Commons Attribution 4.0 International License

followed (with equivalent pauses where a grid would have been) to keep timing consistent, but no memory recall was prompted. Instead, at the end of some low-load trials, a simple yes/no comprehension question about the sentence was asked (e.g., "Did the pirate hide the treasure?") to ensure the participant had listened; these questions were answered correctly at a high rate (>95%), confirming task compliance.

The entire experiment lasted about 40 minutes, including instructions and calibration. After finishing, participants filled out a brief questionnaire on background and language history, and we administered a computerized vocabulary size test as an additional proficiency measure (optional, used to characterize the sample).

Data Analysis

Eye-tracking data were processed using standard procedures in the visual-world paradigm. We defined dynamic Areas of Interest (AOIs) for each of the four object pictures (bounding boxes of approximately 3° radius around each object). Fixations were coded as hitting an AOI if they fell within these bounds. The primary measure of interest was the proportion of trials in which participants fixated the target object (the object eventually mentioned, e.g. the treasure) as the sentence unfolded, especially in the time between the onset of the predictive context and the onset of the target noun. We aligned each trial's time course to the moment of target noun onset (defined by the speech waveform). We then examined looks in the time window roughly from 200 ms before noun onset (to account for any slight lead in prediction) to 800 ms after noun onset (by which time the word should be recognized and fixations converge on the target). Anticipatory looking is typically inferred by an increase in target fixations before the noun is uniquely identified by speech. Because the target's name initially shares phonemes with potential competitors (e.g., if "treasure" and "tree" were both present, both start with "tr-"), we specifically monitored the timeline of when the target started to be preferentially fixated over a phonologically unrelated distractor.

Statistical analysis of the eye-movement data was carried out using growth curve analysis (GCA), a type of multilevel regression suited for time-course data. GCA involves fitting polynomial curves (e.g., linear, quadratic, cubic terms) to the fixation proportion over time for each condition, and comparing these curves between conditions. We binned the time axis into 20 ms intervals and computed the empirical logit of the proportion of target fixations (with an adjustment for 0 and 1 values) at each time slice for each participant and condition. We then fit mixed-effects regression models with orthogonal polynomial time terms (up to third order) as predictors, along with fixed effects of Context (Predictive vs. Control), WM Load (High vs. Low), and their interactions on those time terms. Subject and item were included as random effects (with random intercepts and slopes for critical effects as justified by model comparison). This approach allowed us to assess differences in the shape and timing of fixation curves across conditions. In particular, a significant Context × Load × Time interaction (for instance on the linear or quadratic term) would indicate

This is an open access article under the Creative Commons Attribution 4.0 International License

that the time-course of anticipatory looking in predictive sentences differs between load conditions. Additionally, we included participants' proficiency (standardized test score) as a covariate in an extended model to test for interactions between proficiency and the load effect.

We also analyzed response accuracy in the VSWM task (proportion of correctly recalled grids in High Load trials) to confirm that the high load condition indeed demanded memory resources. For completeness, target identification latency was measured by noting the time of the first fixation on the target after noun onset in each trial, although anticipatory effects are better captured by the pre-noun period as described above. Finally, any trials with tracker loss or blink during the critical interval were excluded (this was < 2% of trials). All analyses were conducted in R, using the lme4 package for mixed models. Significance of fixed effects was assessed via likelihood-ratio tests and by examining 95% confidence intervals around estimates.

Results

Manipulation Checks

Participants performed well on the secondary VSWM task in the High Load condition, but with clear evidence that it was cognitively demanding. Mean recall accuracy for the two-location symmetry span was 78% (SD = 11%), significantly above chance (which would be 0% for random guessing, or 25% if guessing each location independently out of four possibilities per trial). This indicates that participants were actively trying to maintain the visuospatial sequences during the sentences. At the same time, the less-than-ceiling performance confirms that the task taxed their memory (nobody achieved 100%, and many errors were evident in recalling the correct sequence), which is desirable for our load manipulation. In the Low Load condition, where no concurrent memory task was present, the intermittent comprehension questions were answered with 96% accuracy, confirming participants' focus on the sentences even when no extra task was required. There was no evidence of a speed-accuracy tradeoff between the dual task and sentence understanding: accuracy on comprehension questions was uniformly high in both load conditions (when asked), suggesting participants did not entirely sacrifice language processing for the memory task.

Eye-Tracking Measures: Predictive Gaze Timing

We first describe qualitatively the time-course of eye gaze in the critical Predictive versus Control sentence contexts, and how this differed by WM load. Figure 1 (not shown here) would illustrate the proportion of fixations on the target object over time. Under Low Load (no secondary task), participants showed clear anticipatory eye movements in the Predictive condition. Approximately 300–400 ms after the onset of the verb (in sentences like "The pirate hides the treasure…"), the target object (e.g., treasure) began to attract more fixations than the other objects. This divergence occurred well before the target noun was spoken (which typically begins ~the trea- at time 0 by

This is an open access article under the Creative Commons Attribution 4.0 International License

alignment). By about 200 ms before target noun onset, the target was fixated around 40% of the time in predictive trials compared to only ~20% in control trials, indicating that learners were using the verb and scene context to predict the likely referent. This replicates the basic finding of anticipatory processing: L2 listeners, when unburdened by other tasks, can indeed pre-activate a likely upcoming noun and direct their gaze accordingly, much as native speakers do (albeit perhaps not as strongly as reported for natives in prior studies).

In contrast, under High Load (with a concurrent VSWM task), the anticipatory fixation advantage for the target was notably attenuated and temporally delayed. Early in the sentence, predictive and control trials showed overlapping fixation proportions on the target. It was only closer to the noun onset (around 0 ms or even slightly after) that the predictive condition began to pull ahead. In essence, with a visuospatial memory task occupying part of their attention, participants still eventually looked at the target object more if the context was predictive, but this predictability effect emerged roughly 150–200 ms later than in the low load condition. By the time the target word was actually spoken and recognized (around 500 ms after noun onset), both load conditions converged on high target fixations (>70%), meaning all participants identified the referent by the end; however, the timing of the rise in target-directed gaze differed.

Statistical Analysis

The growth curve modeling confirmed these observations. There was a significant main effect of Context Predictability ($\chi^2(1) = 45.2$, $p < 0.001$), such that overall, fixations on the target grew faster and to a higher asymptote in predictive sentences than in non-predictive control sentences (demonstrating the basic predictive gaze effect). Crucially, this effect was qualified by a Context × WM Load interaction on the timing (linear term) of the fixation curve ($p < 0.001$). In the Low Load condition, the target advantage in predictive trials appeared in the pre-nominal time region, whereas in the High Load condition it appeared later, around or just after noun onset. The model's fixed effects estimates indicated that high VSWM load caused a significant slowing of the predictive look onset. Specifically, the interaction term corresponded to an estimated shift of approximately 150 ms in the target fixation curve: under high load the peak of anticipatory gain was delayed ($\beta = -0.22$, SE = 0.05, $p < 0.001$, for the Predictive × HighLoad interaction coefficient). Figure 2 (not shown) would illustrate this by comparing the best-fit curves: the high-load predictive curve lagged behind the low-load predictive curve, though eventually reaching a similar final level of target fixations by ~600 ms after noun onset.

To put it plainly, when participants' visuospatial WM was occupied, they were slower to predict the upcoming referent. They still could predict — the fact that the predictive vs. control difference did appear (and ultimately reached significance even in high load) shows that some anticipatory processing was intact. But they needed more time or more linguistic input before their gaze reflected the prediction. In the low load condition, minimal phonological input ("tre-…") was

This is an open access article under the Creative Commons Attribution 4.0 International License

enough for them to zero in on “treasure” thanks to the supportive context; in the high load condition, their eyes did not move reliably to the treasure until perhaps “treas-” or later, indicating a brief delay in activating the target.

Proficiency Effects

We examined L2 proficiency (as measured by the vocabulary test scores, which ranged from intermediate to advanced) as a factor in the model. Proficiency on its own had a marginal effect on overall target fixation proportions (higher proficiency tended to correlate with slightly faster and stronger target looks across conditions), but more importantly, there was no Proficiency × Load × Context interaction ($\chi^2 < 1$, n.s.). In other words, the high-load delay in prediction was consistent across the proficiency spectrum in our sample. Both mid-level and more advanced learners experienced a similar slowing of anticipatory eye movements under VSWM load. There was a hint that the highest-proficiency participants showed the earliest predictions overall (as one might expect), but even they showed a slow-down when burdened with the secondary task. Thus, we find no evidence that proficiency immunized learners against the effects of VSWM load – even those near-native in English were subject to the capacity constraints imposed by the dual task. This point reinforces the notion that the limitations observed are due to general cognitive load rather than lack of L2 knowledge.

Summary of Key Findings

To summarize, the results confirm that: (1) L2 learners made predictive eye movements to likely upcoming referents in supportive contexts, demonstrating anticipatory processing at the phonological level (before words were fully spoken), and (2) a concurrent visuospatial WM load caused a significant delay (~150 ms) in the onset of these predictive eye movements. The effect of load was robust and statistically significant, while no statistically reliable effect of L2 proficiency on the size of this delay was found. In the next section, we interpret these findings in light of theoretical accounts of WM in language processing and discuss their implications for our understanding of L2 comprehension under dual-task conditions.

Discussion

Our findings provide novel evidence that visuospatial working memory (VSWM) resources play a measurable role in L2 real-time language processing. When L2 listeners were under high visuospatial load, their ability to predict an upcoming word was not eliminated, but it was significantly slower. In this discussion, we unpack the theoretical significance of this VSWM effect, compare it with predictions from capacity-based and interference-based models of WM, and explore implications for multilingual language use and learning (including contexts like captioned video comprehension).

This is an open access article under the Creative Commons Attribution 4.0 International License

VSWM as a Capacity Constraint in L2 Processing

The approximately 150 ms delay in anticipatory fixations under high load suggests that even a task in a different cognitive domain (remembering visual patterns) can sap enough attentional resources to slow language processing. This result aligns well with capacity-based models of working memory. Under the classic capacity view (Just & Carpenter, 1992), working memory has a limited pool of resources for both storage and processing; if part of that pool is occupied by a secondary task, fewer resources remain for language processing operations such as generating predictions. Our experiment demonstrates this principle in action: the visuospatial task competed for central executive attention, thereby delaying lexical prediction in the L2. Importantly, because the secondary task was non-linguistic (involving spatial locations), the interference cannot be attributed to direct competition within the verbal system (e.g., it's not that participants had to rehearse other words that confused them about the upcoming noun). Instead, the effect seems to reflect a domain-general resource bottleneck.

This supports the idea that predictive processing is a resource-intensive, "optional" process that is engaged when sufficient resources are available (Kuperberg & Jaeger, 2016). In low load situations, our L2 listeners proactively used context to anticipate the next word (even before phonological information fully arrived), but under high load, they may have adopted a more wait-and-see strategy, effectively deferring predictive commitments until more bottom-up input came in. The fact that the end-state comprehension did not suffer (everyone understood eventually that the pirate hid the treasure) suggests that prediction is not strictly necessary for comprehension – rather, it is a facilitative process that kicks in given enough cognitive bandwidth, improving processing efficiency by pre-activating likely words. When bandwidth is limited, comprehension shifts to a more reactive mode, and the difference is observed in timing (slower uptake of the input, less anticipatory looking).

Capacity vs. Interference: Insights from the Dual-Task Paradigm

Our results also shed light on the ongoing debate between capacity-based and interference-based accounts of WM in L2 processing (Cunnings, 2017; Juffs & Harrington, 2011). Interference-based models (e.g., Lewis et al., 2006; Cunnings, 2017) argue that the primary limitation of WM is not an all-purpose capacity, but rather the ability to efficiently manage interference between similar items or to retrieve the correct item from memory when needed. From this perspective, L2 processing difficulties (such as less prediction or shallow parsing) might stem from L2 learners encoding or retrieving information in a way that is more susceptible to interference (for instance, weighting irrelevant cues, or not distinguishing similar memory traces sharply enough). How would such a model explain our findings? If predictive processing were hampered solely by interference, one might expect a secondary task to have a large effect only if it introduces confusable content or overlapping representations. A visuospatial load (remembering red-square

This is an open access article under the Creative Commons Attribution 4.0 International License

locations) shares no content or cues with the linguistic task; it neither introduces misleading lexical items nor similar syntactic features. Therefore, a pure interference account might predict little to no impact of such a task on language processing – after all, the retrieval cues for the upcoming noun ("pirate hides the X") are intact and there are no extra competing lexical items in memory (unlike, say, if we had asked participants to remember other English words while listening, which could interfere lexically).

The clear delay we observed under visuospatial load thus leans toward a capacity limitation interpretation: even though the secondary task did not crowd verbal working memory with competing words, it drained some of the general attentional/executive resources needed to generate or commit to a prediction. Our data suggest that L2 predictive processing requires a threshold level of available resources; if diverted elsewhere, prediction is postponed. This does not mean interference plays no role in L2 processing – certainly, other studies show that L2 learners can be lured by incorrect cues or have difficulty ignoring irrelevant information (e.g., considering a discourse-prominent but grammatically inappropriate antecedent, as in Felser & Cunnings, 2012). However, our experiment was not about cue competition, but resource competition, and the outcome underscores that resource competition alone is enough to alter processing behavior.

Interestingly, this capacity view dovetails with the similar effect found by Ito et al. (2018) for a verbal memory load: in both cases, dividing attention slowed prediction in L2 (and L1). It appears that the bottleneck might reside in the central executive's allocation of attention or in the episodic buffer (Baddeley, 2000) which integrates information. When the episodic buffer is preoccupied with holding a visual sequence online, it may have less room to rapidly integrate linguistic context with world knowledge to generate predictions. In terms of Cunnings' proposal (2017), one might reinterpret L2 learners' susceptibility to interference as partly a result of limited capacity to apply cues under time pressure. Our findings suggest that even when L2 learners know how to use cues to predict (as demonstrated in low load), they might not do so if their executive attention is taxed – thereby giving an appearance of "weak prediction" that could alternatively be explained as a strategic adaptation to high cognitive load.

Another theoretical implication concerns the timing of prediction. The ~150 ms delay is modest in absolute terms, but in the realm of rapid online comprehension, it could be consequential. Prior research (e.g., Huettig & Guerra, 2015) has argued that predictive eye movements in native listening often occur in a very tight temporal window (perhaps 200–300 ms before the target word). A delay of 150 ms might mean the difference between having pre-activated a word versus only identifying it as it is spoken. In practical terms, under high load, learners might not benefit from the head-start that prediction provides, potentially rendering their processing effectively more "bottom-up." Over the course of a conversation or a lecture, this could accumulate, possibly contributing to fatigue or reduced understanding in mentally demanding situations.

This is an open access article under the **Creative Commons Attribution 4.0 International License**

No Proficiency Interaction: Universality of WM Effects

One might have expected that more proficient L2 learners, with more automatized language processing, would be less affected by a given cognitive load than less proficient learners. However, our results did not show an interaction between proficiency and VSWM load – the delay in predictive gaze was roughly consistent for intermediate and advanced learners alike. This finding mirrors Linck et al.'s (2014) meta-analytic observation that the WM–performance relationship in L2 remained positive even for highly proficient individuals. It also echoes Cunnings (2017), who noted that differences between L1 and L2 processing persist beyond initial stages and thus likely involve fundamental cognitive processes rather than just lack of knowledge. The lack of a proficiency effect in our study suggests that the constraint we are observing is not something easily overcome by language experience alone. Even those with near-native command of English presumably still have to allocate cognitive resources to parse and predict; when those resources are tied up, they face the same kind of slowdown. This underscores a key point: cognitive constraints are not exclusive to low proficiency. While less proficient learners no doubt have additional difficulties (e.g., smaller vocabulary, less efficient parsing routines), even advanced L2 users – and indeed native speakers – show processing costs under dual-task conditions. Therefore, our results emphasize the universality of WM constraints in language processing, while also reinforcing that L2 users share these basic cognitive limitations. From a theoretical standpoint, it supports models of L2 processing that are not qualitatively distinct from L1 processing but are an outcome of the same cognitive architecture operating under different resource demands (MacDonald, 2013). L2 speakers use the same mechanisms, but often operate closer to capacity, meaning additional loads tip the balance more visibly.

Implications for Multilingual and Multimedia Contexts

In today's globalized and technology-rich environment, L2 and multilingual individuals frequently process language under less-than-ideal cognitive conditions. They might be translating on the fly, watching videos with subtitles, or learning content through a non-native language while juggling visual aids like slides or illustrations. Our findings carry practical implications for such scenarios. The fact that a visuospatial load impedes predictive listening suggests that any multitasking that engages visual memory or attention can hinder language processing efficiency. For example, consider watching a video lecture in an L2 with complex graphics on screen: the viewer's VSWM is engaged in encoding the visual information (graphs, diagrams) which could delay or reduce their ability to predict and integrate the spoken content. This could partially explain why L2 learners often report cognitive overload in multimedia learning settings.

A particularly relevant context is captioned video, where learners listen to L2 audio and read L2 subtitles simultaneously. This is effectively a dual-input situation – auditory and textual – requiring integration of information across modalities. Reading captions likely draws on the visuo-spatial

This is an open access article under the Creative Commons Attribution 4.0 International License

sketchpad (for oculomotor control and visual text processing) and the phonological loop (for subvocal decoding of text), plus executive coordination between reading and listening. Winke, Gass, and Sydorenko (2013) investigated how learners use captions and found considerable individual variation in eye movements: some learners intensely focus on captions, others split attention more evenly. They also identified factors influencing caption use, such as proficiency and possibly memory abilities. Our results suggest that learners with greater VSWM and executive resources might cope better with captions – they can allocate some attention to reading without completely sacrificing listening, thereby still predicting or processing ahead in the audio. In contrast, learners with lower capacity might experience captions as an "additional cognitive load" that slows their processing of the soundtrack. Indeed, Gass et al. (2019) reported that higher WM span learners spent less time fixating captions and still achieved good comprehension, whereas lower span learners often relied on captions more heavily (presumably because processing L2 audio alone was taxing). This aligns with our interpretation that those with more spare resources can distribute attention more flexibly (e.g., glance at captions strategically) while those with fewer resources must devote them either to listening or reading, potentially missing out on predictive processing in the audio if they are busy reading.

For multilingual communicative situations beyond captioning, consider simultaneous interpreters (who listen in one language and speak in another) or bilinguals switching languages in conversation while observing non-verbal cues. Such individuals operate under extreme cognitive load. Our findings, in a modest dual-task scenario, foreshadow the challenges in those contexts: even a relatively simple spatial memory task delayed lexical prediction; how much more might heavy multitasking slow down or disrupt anticipatory processing? It is likely that interpreters and highly proficient bilinguals develop strategies to cope (training to automatize certain processes, using context to chunk information, etc.), but at a cognitive level, they might also rely on exceptional working memory skills. It is telling that professional simultaneous interpreters are often found to have above-average WM capacities (Christoffels et al., 2006), again highlighting that these skills mitigate, though perhaps never fully remove, processing constraints.

Theoretical Contributions and Future Directions

By demonstrating that VSWM load affects L2 predictive eye gaze, our study extends the theoretical framework of WM in SLA in several ways. First, it provides empirical support for Wen's (2012, 2016) call for an "integrated approach" to WM in language learning – one that goes beyond the phonological loop and acknowledges multiple components. Our evidence suggests that the visuospatial sketchpad should be incorporated into models of L2 processing when relevant (e.g., in tasks involving visual input or environments). Second, it contributes to the emerging consensus that L2 processing differences are often a matter of degree rather than kind: under equivalent conditions, skilled L2 listeners can deploy prediction like natives, but they are subject

This is an open access article under the Creative Commons Attribution 4.0 International License

to the same cognitive laws. This argues against any simplistic notion that L2 users categorically do not or cannot predict; instead, it situates L2 predictive processing in a resource-sensitive framework.

Third, our results provide a clear instance of how domain-general cognitive load impacts language processing. This is valuable for theories that seek to connect psycholinguistic behavior to broader cognitive functions. It also encourages interdisciplinary fertilization – for example, cognitive load theory in education (Sweller, 2010) might intersect with our findings when designing instructional materials for L2 learners. If a lesson demands simultaneous visual and auditory processing (as many do), designers should be mindful of not overwhelming learners' VSWM. This could involve, for instance, pacing the introduction of visual information, using visuals that directly support the auditory message (thereby perhaps reducing net load by providing complementary cues), or training learners to strategically allocate attention.

Finally, we consider future research directions. While our study focused on the phonological level of prediction (listeners anticipating a word form given context), future work could examine other levels: semantic prediction (anticipating meaning without knowing exact form) or syntactic prediction (anticipating a structure or grammatical feature). It would be informative to see if VSWM load similarly affects those. For example, does a spatial memory task also delay the use of syntactic cues (like case markers or word order patterns) in prediction? Additionally, while our high load was visuo-spatial, one could test other types of load (e.g., an executive control task like an N-back on shapes, or a secondary motor task) to further map out which resources are critical for prediction. The fact that proficiency did not interact with load in our data might be revisited with a more diverse sample: perhaps at very low proficiency, prediction is minimal regardless of load (due to lack of linguistic knowledge), whereas at high proficiency we saw robust prediction that was uniformly load-sensitive. A curvilinear relationship might exist across the entire spectrum of proficiency and load, which could be explored with beginners or near-natives specifically.

Another interesting extension would be to use neurophysiological measures (ERPs) alongside eye-tracking to see how load affects brain signatures of prediction in L2. Would the well-known ERP correlates of prediction (such as pre-nominal positivities or reduced N400s for expected words) be diminished under VSWM load? Such data could corroborate our gaze findings and provide deeper insight into whether the delay is due to postponing prediction or simply making weaker predictions that only solidify upon hearing more input.

In summary, our discussion highlights that VSWM plays a role in L2 predictive processing by serving as a necessary resource for timely anticipation. The study's outcomes favor a capacity-based interpretation of WM effects, demonstrate consistency across proficiency levels, and carry implications for real-world L2 usage, particularly in multimedia learning environments. We now conclude by recapping the main contributions and practical takeaways of this research.

This is an open access article under the **Creative Commons Attribution 4.0 International License**

Conclusion

This study investigated how a visuospatial working memory load influences L2 learners' predictive eye gaze during spoken word recognition. The key finding was that imposing a concurrent VSWM task (symmetry span) caused a significant ~150 ms delay in learners' anticipatory fixations to contextually predictable referents, relative to no-load conditions. In essence, L2 listeners still made predictions about upcoming words, but those predictions were slower when part of their attention was diverted to a visual memory task. Notably, this effect held true across proficiency levels, indicating that even advanced L2 users are subject to capacity-based constraints on processing speed. Theoretically, these results extend WM models in SLA (e.g., Wen, 2012) by demonstrating a functional role for the oft-neglected visuospatial sketchpad in language processing. They support the view that predictive processing in an L2 is resource-dependent and can be hindered by domain-general cognitive load, rather than an all-or-nothing ability tied solely to proficiency. Practically, our findings suggest that L2 educators and learners should be mindful of cognitive load in multimedia or dual-task situations – for example, the use of captions, images, or simultaneous activities – as these can impact the efficiency of real-time comprehension. By illuminating how VSWM load affects L2 predictive mechanisms, this study contributes to a more comprehensive understanding of the interplay between memory and language in bilingual minds, and encourages the incorporation of multimodal cognitive factors into second language processing research and pedagogy.

References

Alptekin, C., & Erçetin, G. (2010). The role of L1 and L2 working memory in literal and inferential comprehension in L2 reading. Journal of research in reading, 33(2), 206-219.

Berg, D. H. (2008). Working memory and arithmetic calculation in children: The contributory roles of processing speed, short-term memory, and reading. Journal of experimental child psychology, 99(4), 288-308.

Cunnings, I. (2017). Parsing and working memory in bilingual sentence processing. Bilingualism: Language and Cognition, 20(4), 659-678.

Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. Journal of Memory and Language, 66(1), 226-248.

Gass, S., Winke, P., Isbell, D. R., & Ahn, J. (2019). How captions help people learn languages: A working-memory, eye-tracking study.

Giofrè, D., Mammarella, I. C., Ronconi, L., & Cornoldi, C. (2013). Visuospatial working memory in intuitive geometry, and in academic achievement in geometry. Learning and Individual Differences, 23, 114-122.

This is an open access article under the **Creative Commons Attribution 4.0 International License**

Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. Studies in second language acquisition, 34(3), 445-474.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. Studies in second language acquisition, 14(1), 25-38.

Hvelplund, K. T. (2011). Allocation of cognitive resources in translation: An eye-tracking and key-logging study. Frederiksberg: Copenhagen Business School (CBS).

Ito, A., & Knoeferle, P. (2023). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. Behavior Research Methods, 55(7), 3461-3493.

Ito, A., Corley, M., & Pickering, M. J. (2018). A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension. Bilingualism: Language and Cognition, 21(2), 251-264.

Juffs, A. (2004). Representation, processing and working memory in a second language. Transactions of the Philological Society, 102(2), 199-225.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. Language teaching, 44(2), 137-166.

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different?. Linguistic Approaches to Bilingualism, 4(2), 257-282.

Kim, Y., Payant, C., & Pearson, P. (2015). The intersection of task-based interaction, task complexity, and working memory: L2 question development through recasts in a laboratory setting. Studies in Second Language Acquisition, 37(3), 549-581.

Leeser, M. J. (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: Topic familiarity and working memory. Language learning, 57(2), 229-270.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. Psychonomic bulletin & review, 21(4), 861-883.

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.

Monnier, C., Boiché, J., Armandon, P., Baudoin, S., & Bellocchi, S. (2022). Is bilingualism associated with better working memory capacity? A meta-analysis. International Journal of Bilingual Education and Bilingualism, 25(6), 2229-2255.

This is an open access article under the **Creative Commons Attribution 4.0 International License**

Palladino, P., & Cornoldi, C. (2004). Working memory performance of Italian students with foreign language learning difficulties. Learning and individual differences, 14(3), 137-151.

Roberts, L., & Siyanova-Chanturia, A. (2013). Using eye-tracking to investigate topics in L2 acquisition and L2 processing. Studies in Second Language Acquisition, 35(2), 213-235.

Rönnberg, J., Lunner, T., Ng, E. H. N., Lidestam, B., Zekveld, A. A., Sörqvist, P., ... & Stenfelt, S. (2016). Hearing impairment, cognition and speech understanding: exploratory factor analyses of a comprehensive test battery for a group of hearing aid users, the n200 study. International Journal of Audiology, 55(11), 623-642.

Smith, M. S., & Truscott, J. (2014). The multilingual mind: A modular processing perspective. Cambridge University Press.

Van den Noort, M. W., Bosch, P., & Hugdahl, K. (2006). Foreign language proficiency and working memory capacity. European Psychologist, 11(4), 289-296.

Wen, Z. (2012). Working memory and second language learning. International Journal of Applied Linguistics, 22(1), 1-22.

Winke, P., Gass, S., & Sydorenko, T. (2013). Factors influencing the use of captions by foreign language learners: An eye-tracking study. The Modern language journal, 97(1), 254-275.

Received: 25.11.2025
Revised: 10.12.2025
Accepted: 18.01.2026
Published: 21.01.2026

This is an open access article under the **Creative Commons Attribution 4.0 International License**