# *Working Memory and Second Language Oral Production: A Five-Year Systematic Review*

**Shiyu Li[1,a,*]**

[1]*Department of General and Typological Linguistics, Ludwig Maximilian University of Munich, Geschwister-Scholl-Platz 1, Munich, Germany*
*a. shiyu.liregen@gmail.com*
*\*corresponding author*

*Abstract:* This article presents a comprehensive synthesis of the literature exploring the role of working memory in second language (L2) oral production. It systematically reviewed nine studies investigating the associations between working memory and L2 speaking, focusing on research foci, working memory measurement, speaking tasks, and performance. The synthesized results revealed a positive correlation between working memory and language production aspects, including accuracy, lexical complexity, and fluency. Learners with stronger working memory demonstrated superior performance, particularly in complex tasks. However, some studies did not find a significant predictive relationship, possibly due to sample homogeneity, sampling heterogeneity, or task complexity. This systematic review sheds light on the crucial role of working memory in L2 oral production and provides valuable insights for language educators and researchers.

*Keywords:* working memory, second language oral production, systematic review

## 1. Introduction

Oral production, as a fundamental language skill, holds paramount importance in the realm of L2 acquisition. It serves as a crucial means of communication, enabling learners to express themselves fluently and accurately in their target language. However, the process of generating coherent and articulate speech involves complex cognitive mechanisms. One such mechanism that garnered substantial attention is working memory, a cognitive process responsible for the temporary storage and manipulation of information during ongoing cognitive tasks [1]. Particularly in the initial stages of L2 learning, speaking often entails effortful retrieval, selection, and articulation of linguistic items, underscoring the paramount significance of working memory in L2 speaking [2]. Exploring the association between working memory and L2 oral production bears profound pedagogical implications within the domain of language teaching. Understanding how working memory influences oral production can inform language educators about effective strategies to enhance learners' speaking proficiency. However, despite the pedagogical importance of this association, there has been a limited number of literature reviews on this specific topic. Existing literature reviewed either narrowly focused on L2 oral production in the context of translation and interpreting in relation to working memory [3], or took a broader perspective, covering multiple language skills such as reading, speaking, lexical development, etc., but neglecting detailed

examinations of oral production [4]. This leaves a notable gap in understanding the impact of working memory on L2 oral production in various contexts. Therefore, the present systematic review aims to bridge this gap by addressing the following research questions:

1. What were the specific research foci in the selected studies?

2. What were the characteristics of the participants (samples) involved?

3. What instruments were used to measure working memory?

4. What types of speaking tasks were utilized, and how was L2 speaking performance measured?

5. What were the reported findings regarding the influence of working memory on L2 oral production?

## 2. Methods

## 2.1. Search Strategy

The search strategy employed for this systematic review involved a comprehensive exploration of major databases in psychology, linguistics, and education, ensuring a comprehensive coverage of relevant literature. The selected databases included ERIC, LLBA, PsycInfo, and the general database Web of Science Core Collection. The search was conducted in June 2023, utilizing a Boolean search query: ("Working Memory" OR "Short-term Memory" OR "Working Memory Capacity" OR "Cognitive Load" OR "Memory Capacity") AND ("Second Language" OR "L2" OR "Foreign Language" OR "Additional Language" OR "ESL" OR "EFL") AND ("Oral Production" OR "Spoken Language" OR "Speaking Performance" OR "Speech Production" OR "Oral Proficiency").

## 2.2. Literature Selection

To ensure the credibility and reliability of the sources, the search results were filtered to include only peer-reviewed articles published in English within the specified timeframe from 2018 to 2023. The establishment of clear inclusion and exclusion criteria, as depicted in Table 1, provided a systematic framework for the selection process, enabling a meticulous evaluation of each paper's eligibility for analysis.

Table 1: Inclusion and exclusion criteria.

| Criteria | Inclusion Criteria | Exclusion Criteria |
|---|---|---|
| Publication Type | Peer-reviewed journal papers | Non-peer-reviewed sources, conference abstracts, dissertations |
| Language | Written in English | Written in languages other than English |
| Research Focus | Studies investigating the relationship between working memory and L2 oral production | Studies not directly related to working memory and L2 oral production |
| Study Design | Empirical studies | Theoretical papers, reviews, case studies |
| Full-Text Availability | Papers with full-text available | Papers without full-text availability |

Table 1: (continued).

| Publication Date | Since 2018 | Before 2018 |
|---|---|---|

Based on the established inclusion and exclusion criteria, a systematic review following the PRISMA guidelines (Figure 1) was conducted. After several rounds of screening, nine papers that met the inclusion criteria were eventually retained and labeled sequentially as ID1-ID9, as shown in Table 2 and Table 3, conveniently located within the appendix for easy reference.
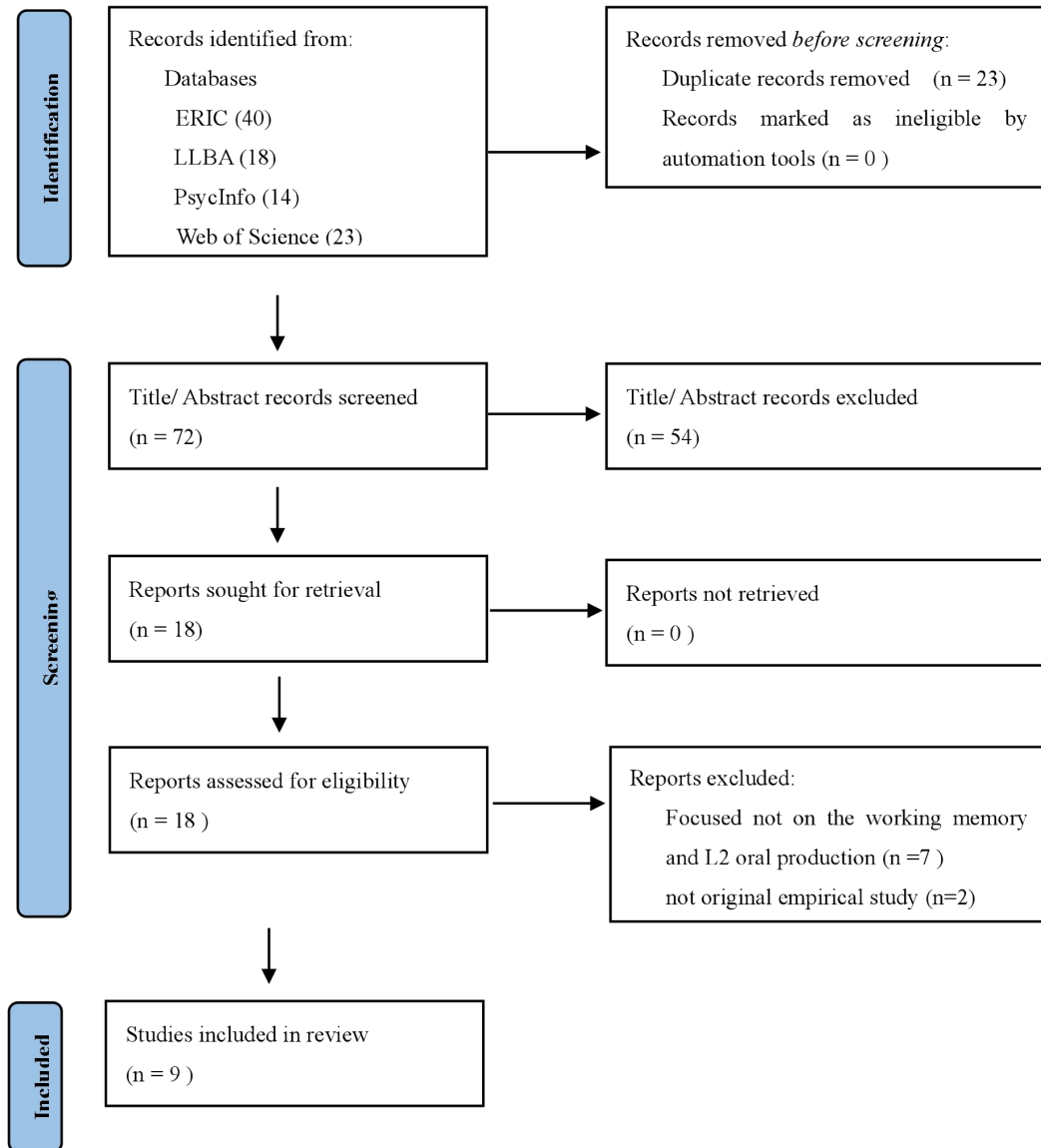


Figure 1: Literature selection process.

## 3.    Results

## 3.1.    Research Foci

Among the selected studies, the examination of working memory's role in L2 oral production varied. One study (ID7) exclusively focused on working memory as a singular influencing factor. In

five studies, working memory was considered as one of several factors impacting oral production, alongside variables such as task complexity (ID4, ID9), modality (ID9), linguistic variables (ID1), language ability (ID1, ID5), and pre-task explicit instruction (ID8). Additionally, in three studies (ID2, ID3, ID6), working memory was investigated as a mediator between different predictors and L2 speaking performance.

## 3.2. Sample Characteristics

The reviewed studies primarily involved university students as participants (7 out of 9 studies, including ID1, ID3, ID4, ID5, ID6, ID7, and ID9). One study (ID2) specifically focused on secondary school students, while another study (ID8) included learners at the A2 proficiency level from five different language institutes.

Regarding the target languages, there was relative homogeneity across the 9 studies. Six studies (ID1, ID2, ID3, ID4, ID8, and ID9) focused on English as the target language. Additionally, individual studies explored the acquisition of Chinese, French, and Spanish as target languages. The participants' native languages (L1s) in this dataset were diverse, including Chinese, English, Korean, Philippine, Iranian, Vietnamese, Samoan, Russian, Arabic, French, or mixed in the case of a sample consisting of international students with varied L1 backgrounds.

## 3.3. Measurement of Working Memory

Working memory measurement can be divided into two categories: verbal and nonverbal, depending on whether language-based stimuli are used. Verbal working memory measures involve processing linguistic material, such as words or sentences, while nonverbal measures require processing nonlinguistic material, such as numeric digits, math equations, or visuospatial images [13]–[15] Nonverbal measures employed in the selected studies included operation span (ID3, ID6, ID8, ID9) and digit span tasks (ID2, ID3, ID7). Verbal tests, on the other hand, utilized linguistic stimuli, such as reading span (ID4, ID9), letter span (ID1), and nonword span (ID5).

Regarding language use in the working memory tests, four studies (ID3, ID4, ID7, ID9) employed participants' L1. Two studies (ID1, ID8) utilized their L2, while one study (ID2) included both L1 and L2 tasks. In two studies (ID5, ID6) with participants from diverse language backgrounds, the operation span test was conducted in English, possibly chosen as a neutral language for assessment.

## 3.4. Speaking Tasks and Measures

The selected studies encompassed two main categories of speaking tasks: eliciting tasks and independent speaking tasks. Eliciting tasks, present in 5 out of the 9 studies, aimed to stimulate spoken language output from participants [14], [16]. These tasks included picture description tests (ID1, ID4) and video or picture retelling tasks (ID2, ID6, ID8). On the other hand, independent speaking tasks, featured in the remaining 4 studies, involved participants speaking without specific prompts, such as narrative topic telling tasks (ID3, ID5), reading-aloud tasks (ID7), picture-based narration tasks (ID7), and monologic tasks (ID9).

The implementation of these speaking tasks exhibited variations in terms of speaking time, availability of planning time, and note-taking. Only 3 studies (ID1, ID3, ID5) reported the allotted time for participants to complete the speaking tasks, ranging from 45 seconds to three minutes. Additionally, 4 studies (ID1, ID3, ID5, ID4) allowed participants planning time, which varied between 30 seconds and three minutes. Two studies (ID3, ID7) permitted participants to take notes during the task.

Regarding the measured constructs, 6 studies examined complexity, accuracy, and fluency (CAF) together (ID2, ID4, ID5, ID6, ID8, ID9); one investigated lexical complexity and fluency (ID3); two studies employed subjective ratings, focusing on the comprehensibility (ID1) and pronunciation and proficiency (ID7).

## 3.5. Related Findings

Five studies found an association between working memory and second language speaking performance. Study (ID2) demonstrated that learners with stronger working memory exhibited higher accuracy and lexical complexity in L2 speech tasks, particularly those involving intentional reasoning. Additionally, Study (ID4) revealed that working memory capacity significantly influenced L2 speaking performance, especially in cognitively demanding tasks, highlighting the advantages for learners with greater working memory capacity in managing linguistic aspects and task completion simultaneously. Regarding L2 fluency, Study (ID3) found a positive correlation between backward-span task scores and learners' articulation rates during task repetition, indicating the predictive power of working memory for fluency. Similarly, Study (ID8) showed that working memory strongly predicted various language production aspects, including relativization rate, complexity, and fluency. Moreover, Study (ID6) revealed significant correlations between working memory and accuracy and fluency measures in unpressured within-task planners, while no significant correlations were observed in strategic planners. These findings underscored the dynamic nature of working memory functions, which could be activated or inhibited depending on task conditions.

## 4. Discussion

## 4.1. Methodology Issues

### 4.1.1. Sampling

Sampling issues in research merit careful consideration due to their significant impact on study outcomes. Researchers must attend to crucial sample characteristics, including age, learning experience, stage, and proficiency level, when selecting participants. Analyzing and consulting these characteristics is vital for accurate interpretation of research results. Biased sampling is another concern arising from research synthesis, where the selected sample may not represent the entire learner population. For example, in Study (ID4), the exclusive focus on English education majors in a graduate program in South Korea may limit the generalizability of the findings to a broader population of English learners. Sampling heterogeneity is a critical consideration as well, and failure to address it can introduce confounding variables, undermining study validity. For instance, in Study (ID5), the sample included both native English speakers and non-native English learners, necessitating control for language background during data analysis to ensure the robustness of the results. Overall, researchers should exercise diligence in addressing sampling issues to enhance the reliability and applicability of their findings.

### 4.1.2. Working Memory Measurement

Regarding working memory measurement, it is crucial to consider the potential influence of L2 proficiency on L2 working memory tests, which could introduce confounding factors. To address this concern, it is advisable to use L1-based or language-neutral stimuli instead of L2 working

memory tests. However, it is worth noting that two of the nine primary studies included in this review used L2 working memory tests, which might have impacted the results to some extent. Furthermore, upon analyzing the data set, it is evident that nonverbal tests were prevalent in studies ID2, ID3, ID6, ID7, ID8, and ID9. Some studies suggested that nonverbal working memory tests may be less influenced by language background and educational experience, making them more effective for diverse research groups [17], [18]. However, other research indicated that language-based working memory measurement may be more closely related to key cognitive and learning outcomes, such as reading and vocabulary learning [19]. Therefore, the choice of testing should depend on the specific research goals and the target population.

### 4.1.3. Speaking Task and Measurement

The measurement of speaking performance involved three main components: the selection and design of the speaking task, its execution, and the scoring process. However, many studies lacked clear rationales for task selection and detailed task implementation, such as pre-task planning time, time constraints, and note-taking permissions. This lack of transparency may affect the validity and generalizability of the findings. Regarding scoring methods, some studies used subjective ratings, while the majority employed CAF measures. The objectivity of CAF measures has made them popular in L2 speaking research, but subjective ratings may capture additional aspects like content and organization. Therefore, combining these two methods is recommended for a more comprehensive evaluation of results.

### 4.2. Issues on the Related Findings

In four studies within the review, the relationship between working memory and L2 speaking performance was not observed. Several factors may account for these findings. Firstly, high sample homogeneity could have contributed to the lack of hypothesized results. For instance, in study (ID9), all EFL learners had similar educational backgrounds in English language and literature, making it difficult to draw differentiated conclusions. Secondly, sampling heterogeneity could also have influenced the outcomes. For example, study (ID5) included both native English speakers and non-native English speakers as participants, while study (ID7) involved Chinese French learners with varying degrees of English language learning experiences. These factors might not have been adequately considered, leading to potential biases in the results. Thirdly, the simplicity and execution of the speaking tasks may influence the outcomes. In simpler tasks, such as the Independent oral production task in study (ID5), where participants were asked to describe their typical weekday activities during the school year, the cognitive load during speaking might have been minimal. As a result, the participants may not have heavily relied on their working memory during the task. Instead, their performance in these tasks could have been predominantly influenced by their language-related skills and proficiency.

### 4.3. Suggestions for Future Research

To improve the relevance of the study, researchers should carefully select participants, considering factors such as age, learning experience, proficiency level, and learning stage. Avoiding biased sampling and addressing sampling heterogeneity are also essential. For accurate assessments, using language-neutral or L1-based stimuli is recommended when measuring working memory, while aligning the choice of working memory testing with research goals and the target population. To enhance the validity and generalizability of speaking performance measurements, researchers should provide transparent task selection and implementation details. Combining subjective rating

methods with CAF measures can be beneficial. In cases where a significant relationship between working memory and L2 speaking is not observed, potential reasons may include sample homogeneity, sampling heterogeneity, or task complexity. To gain a comprehensive understanding of working memory's role in L2 speaking performance, future research should address sampling issues and carefully control for task complexities.

## 4.4. Limitation

The current research also had limitations. Firstly, the review focused solely on journal articles from reputable databases to prioritize quality, omitting conference papers and theses. Secondly, citation or reference network analysis was not incorporated to maintain focus and conciseness. For future research, a broader scope could be considered.

## 5. Conclusion

This systematic review delves into the relationship between working memory and L2 oral production within the scope of research conducted from 2018 to 2023. It sheds crucial light on the pivotal role of working memory in L2 oral production, offering valuable insights for language educators and researchers. The comprehensive analysis encompasses various aspects, including research focus, sample characteristics, working memory measurement, speaking tasks, performance measurement, and related findings. Among the findings, several studies revealed a positive correlation between working memory and language production aspects, such as accuracy, lexical complexity, and fluency. Stronger working memory was associated with better performance, particularly in complex tasks. However, it's worth noting that not all studies reported a significant predictive relationship, possibly influenced by factors such as sample homogeneity, sampling heterogeneity, or task complexity. For future research endeavors, researchers should consider these factors during research design and implementation to further investigate and refine the relationship between working memory and L2 oral production. Such efforts will contribute to a deeper understanding of this significant aspect of language acquisition.

## References

[1]  S. Li, "Working Memory and Second Language Learning," in The Routledge Handbook of Second Language Acquisition and Psycholinguistics, 1st ed.New York: Routledge, 2022, pp. 348–360. doi: 10.4324/9781003018872-32.

[2]  S. Li and M. Fu, "Strategic and Unpressured Within-Task Planning and Their Associations with Working Memory," Lang. Teach. Res., vol. 22, no. 2, pp. 230–253, Mar. 2018, doi: 10.1177/1362168816684367.

[3]  H. Wang, Y. Zheng, and H. Xu, "An Overview of the Application of Working Memory in Interpreting Research," Foreign Languages Research, vol. 34, no. 5, pp. 70–74, 2017, doi: 10.13978/j.cnki.wyyj.2017.05.014.

[4]  A. Juffs and M. Harrington, "Aspects of working memory in L2 learning," Lang. Teach., vol. 44, no. 2, pp. 137–166, Apr. 2011, doi: 10.1017/S0261444810000509.

[5]  O.-S. Lee and J. Park, "Understanding L2 Speech Production: Implications for Teaching Speaking in EFL Classroom," J. Asia TEFL, vol. 17, no. 3, pp. 808–823, Autumn 2020.

[6]  A. Awwad and P. Tavakoli, "Task complexity, language proficiency and working memory: Interaction effects on second language speech performance," IRAL Int. Rev. Appl. Linguist. Lang. Teach., vol. 60, no. 2, pp. 169–196, Jun. 2022, doi: 10.1515/iral-2018-0378.

[7]  Van Lang University, Vietnam, T. P. Duong, V. H. H. Le, and Van Lang University, Vietnam, "How Working Memory and Prior Vocabulary Knowledge Influence the Impact of Task Repetition on L2 Oral Performance: Insights into Vietnamese EFL Learners," Teach. Engl. Second Foreign Lang. J.--TESL-EJ, vol. 26, no. 3, Nov. 2022, doi: 10.55593/ej.26103a12.

[8]     Chung-Ang University, J.-E. Hyun, and J.-H. Lee, "The Effects of Task Complexity and Working Memory on Korean Adult Learners' English Speaking Performance," Engl. Teach., vol. 73, no. 1, pp. 115–134, Mar. 2018, doi: 10.15858/engtea.73.1.201803.115.

[9]     H. I. Park, M. Solon, C. Henderson, and M. Dehghan-Chaleshtori, "The Roles of Working Memory and Oral Language Abilities in Elicited Imitation Performance," Mod. Lang. J., vol. 104, no. 1, pp. 133–151, Mar. 2020, doi: 10.1111/modl.12618.

[10]    D. Simard, T. Molokopeeva, and Y. Q. Zhang, "The Contribution of Working Memory to L2 French Pronunciation among Adult Language Learners," Can. Mod. Lang. Rev., vol. 76, no. 1, pp. 50–69, Feb. 2020, doi: 10.3138/cmlr.2018-0223.

[11]    M. Sadeghi and M. Pourhaji, "The contributions of working memory and pre-task explicit instruction to L2 oral performance," System, vol. 96, p. 102409, Feb. 2021, doi: 10.1016/j.system.2020.102409.

[12]    M. Cho, "Task complexity, modality, and working memory in L2 task performance," System, vol. 72, pp. 85–98, Feb. 2018, doi: 10.1016/j.system.2017.10.010.

[13]    M. Daneman and P. M. Merikle, "Working memory and language comprehension: A meta-analysis," Psychon. Bull. Rev., vol. 3, no. 4, pp. 422–433, Dec. 1996, doi: 10.3758/BF03214546.

[14]    S. Li, "Working memory and second language writing: A systematic review," Stud. Second Lang. Acquis., pp. 1–33, Apr. 2023, doi: 10.1017/S0272263123000189.

[15]    J. A. Linck, P. Osthus, J. T. Koeth, and M. F. Bunting, "Working memory and second language comprehension and production: A meta-analysis," Psychon. Bull. Rev., vol. 21, no. 4, pp. 861–883, Aug. 2014, doi: 10.3758/s13423-013-0565-2.

[16]    A. Amari, Department of English, Faculty of Humanities, Abadan Branch, Islamic Azad University, Abadan, Iran, B. Gorjian, and Department of English, Faculty of Humanities, Abadan Branch, Islamic Azad University, Abadan, Iran, "The Effect of Eliciting Tasks on Listening Comprehension among Iranian Pre-Intermediate EFL Learners," Int. J. Res. Engl. Educ., vol. 4, no. 2, pp. 42–60, Jun. 2019, doi: 10.29252/ijree.4.2.42.

[17]    T. P. Alloway, S. E. Gathercole, and S. J. Pickering, "Verbal and Visuospatial Short-Term and Working Memory in Children: Are They Separable?," Child Dev., vol. 77, no. 6, pp. 1698–1716, Nov. 2006, doi: 10.1111/j.1467-8624.2006.00968.x.

[18]    P. M. J. Engel De Abreu and S. E. Gathercole, "Executive and phonological processes in second-language acquisition.," J. Educ. Psychol., vol. 104, no. 4, pp. 974–986, Nov. 2012, doi: 10.1037/a0028390.

[19]    A. Baddeley, S. Gathercole, and C. Papagno, "The phonological loop as a language learning device.," Psychol. Rev., vol. 105, no. 1, pp. 158–173, 1998, doi: 10.1037/0033-295X.105.1.158.

## Appendix

Table 2: Research focus and sample characteristics of selected studies.

| ID | Reference | Foci | Sample |
|---|---|---|---|
| ID1 | [5] | Relationship between 5 potential predictor variables for L2 speech production | 51 University Korean EFL learners |
| ID2 | [6] | Effects of task complexity on L2 speech performance; mediating role of language proficiency and working memory | 48 male Arabic EFL learners at a private secondary school |
| ID3 | [7] | Task repetition effects on L2 speech performance; moderating role of prior vocabulary knowledge and working memory | 40 Vietnamese third-year university EFL students |
| ID4 | [8] | Task complexity and working memory effects on L2 speaking performance | 20 Korean EFL learners majored in English education. |

Table 2: (continued).

| ID5 | [9] | The role of working memory and oral language abilities in L2 elicited imitation performance | 78 Spanish learners (74 L1 English, 4 L1 not English) |
|-----|-----|------|------|
| ID6 | [2] | Strategic and unpressured within-task planning effects on L2 oral production; working memory mediating role. | 29 University Chinese learners with diverse L1s |
| ID7 | [10] | Working memory's contribution to L2 pronunciation and proficiency | 30 Chinese French learners |
| ID8 | [11] | Contributions of working memory and pre-task explicit instruction to L2 oral performance | 103 Iran EFL learners with A2 level |
| ID9 | [12] | Task complexity, modality effects on performance; working memory's mediating role | 39 University Korean EFL students |

Table 3: Methodology and results of selected studies.

| ID | Working Memory test | Speaking Task/ Measurement | Related findings |
|----|------|------|------|
| ID1 | Forward letter span test | Picture-description test/ Comprehensibility | Short-term memory was related to speech performance. |
| ID2 | Backward-digit span tests | Vedio retelling task/ CAF | Language proficiency and working memory were predictors of speech performance, with working memory specifically affecting accuracy and lexical complexity. |
| ID3 | Backward-digit span task; operation span task | Narrative topic telling task/ lexical complexity and fluency | Working memory may be a predictor of fluency in L2 speaking. |
| ID4 | Reading span task | Picture description tasks (simple and complex)/ CAF | There was a significant relationship between working memory capacity and L2 speaking performance, especially in complex tasks. |
| ID5 | Nonword repetition task | Narrative topic telling task/ CAF | Oral language abilities predicted L2 performance, while working memory did not. |
| ID6 | Operation span test | Vedio retelling task/ CAF | Working memory played a more mediating role in unpressured within-task planning compared to strategic planning. |

Table 3: (continued).

| ID7 | Highest number task | Read-aloud task; picture-based narration task/ Individual rating | Working memory had an effect on L2 pronunciation accuracy. |
|---|---|---|---|
| ID8 | Operation span task | Picture story-retelling task/ CAF | Working memory positively predicted global complexity, fluency, and accurate and frequent relativization in speech. |
| ID9 | Reading span; operation span tests | Monologic tasks (simple and complex)/ CAF | Working memory showed no relationship with task performance |