**PRISMA flow diagram for systematic review**

No. of records identified through database searching (e.g., PsycINFO, ProQuest) (*n* = 845)

No. of duplicates removed (*n* = 110)

**Identification**

No. of additional records identified through references (*n* = 4)

No. of records screening based on the titles, abstracts, and methods

(*n* = 739)

No. of records excluded

(*n* =389)

**Screening**

No. of full-text articles not meeting inclusion criteria (*n* = 329)

No. of full-text records assessed for eligibility (*n* =350)

Studies included in review

(*n* = 21)

**Included**

**Coding Protocol**

|  |  |  |
| --- | --- | --- |
| Features | | Descriptors |
| **Study identification** | |  |
|  | Author | The author(s) of the publication |
|  | Title | Title of the publication |
|  | Year | Year that journal article was published, or thesis defended |
|  | Publication Type | Type of the study that was published (i.e., journal, PhD thesis) |
| **Learner variables** | |  |
|  | L1 | The first language of the participants |
|  | L2 | Target language in which participants encountered |
|  | Context | Foreign or second language learning environment (i.e., FL, SL) |
|  | Country/Region | The country where the research was conducted |
|  | Script | Whether L1 and L2 share the same script (i.e., yes, no, mixed) |
|  | Institutional level | Elementary school, secondary school, high school, university, or language institute |
| **Treatment variables** | |  |
|  | Mode of input | Listening, reading, mixed, or unreported |
|  | Material type | Academic discourse, non-academic discourse, mixed, or not reported |
| **Notetaking features** | | |
|  | Notetaking behavior | Allowed notetaking, required notetaking, or not reported |
|  | Num\_notetaking\_sessions | Number of notetaking sessions |
|  | Provision of notetaking instruction | Presence, or absence |
|  | Instruction\_length | Total length of instruction time in minutes |
|  | Type of notetaking strategy | Linear learning strategy, generative learning strategy, or both |
|  | Opportunity of review notes | Yes, no, or not reported |
| **Outcome features** | |  |
|  | Target learning | Linguistic forms, listening comprehension, reading comprehension, writing |
|  | Test max score | Maximum score for the test |
|  | N | The total number of participants |
|  | n | Participant number in each group |
|  | Post\_mean | Learners’ learning outcome in the immediate posttest |
|  | Post\_SD | The standard deviation of the posttest scores |
| **Methodology** | |  |
| Measurement type | | Recognition, recall, writing, or not reported |
| Testing instrument | | Meaning recall, form recall, multiple-choice, T/F question, etc. |
| **Study quality** | | Report whether the study reported some aspects of study quality such as pretest, instrument reliability, etc. |

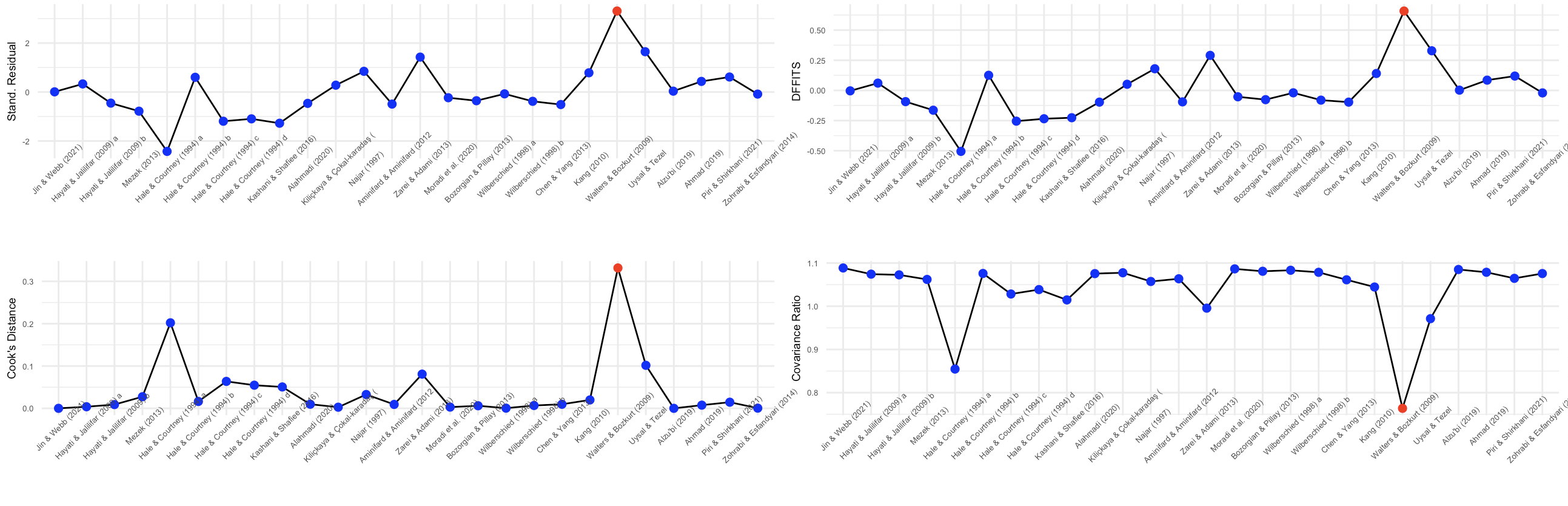
**Sensitivity Analyses**

Following previous meta-analyses (e.g., Kim & Webb, 2022; Shintani et al., 2013), we first converted the effect sizes into z scores, and did not find any absolute value that was larger than 2.0. We then used the *meta* package in R software to identify outliers. As Harrer et al. (2021) suggested, this package provides measures such as DFFITS values, Cook’s distances, and Covariance ratio (see Harrer et al., 2021, for their definitions and equations) to detect outliers. Plots in Figure 1 revealed that one study (Walters & Bozkurt, 2009), which is marked with a red circle, is considered as a potential outlier. And subsequent analyses in second research question also identified it as the only potential outlier. However, because each study was independently conducted and included a different group of students and varying learning conditions, studies identified as outliers do not necessarily mean the study is an outlier that does not reflect normal language learning. Therefore, it is not appropriate to simply delete “outlier” studies from the analysis (see, e.g., Hunter & Schmidt, 2004, for arguments about how to treat outliers). We followed Viechtbauer and Cheung’s (2010) guidance and reran the whole analysis while excluding the studies identified as influential and compared the results to the results obtained when including all studies. The similar pattern in the magnitude of effect sizes revealed that the analyses could be interpreted as robust.

We also examined the potential influence of publication bias—studies finding significant or large effects tend to be published—on the current data. Checking this bias is a critical procedure to examine the credibility of the overall effect sizes. We inspected the funnel plot (Figure 2) and conducted Egger’s test. The funnel plot indicated no apparent publication bias. Egger’s regression test (*t* = 1.84, *p* = .079) also ﻿indicated no substantial influences from publication bias on the findings.

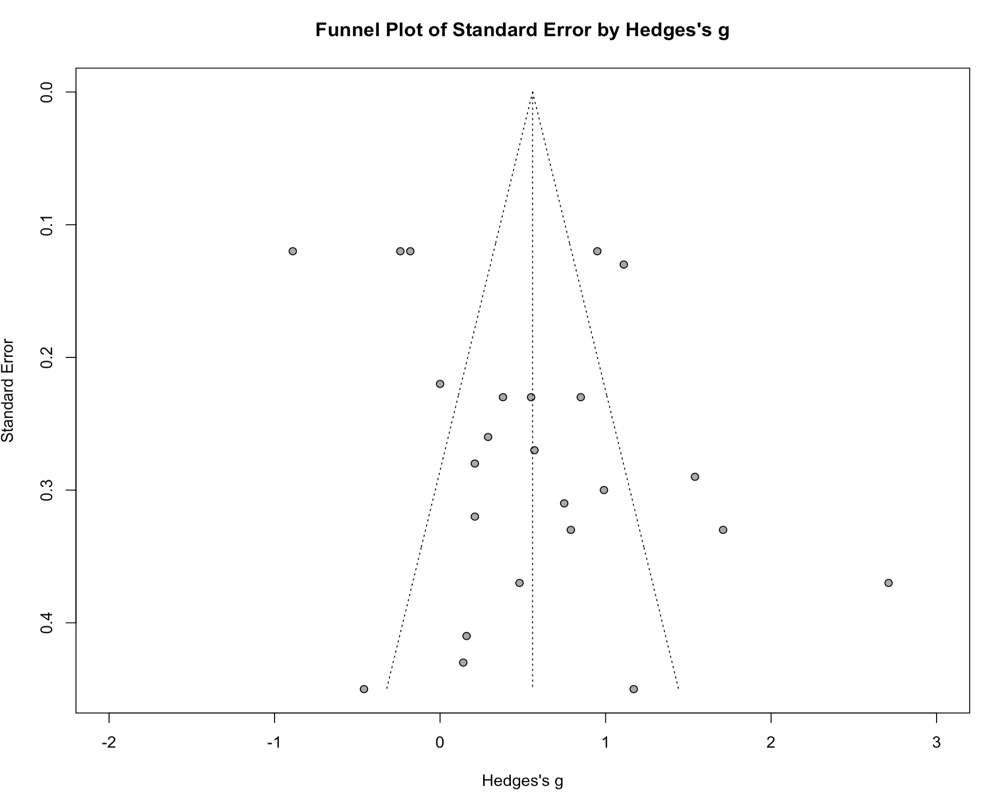
**Figure 1**

*﻿Plot of the (a) studentized deleted residuals, (b) Cook’s distances, (c) DFFITS values, and (d) COVRATIO values for 19 effect sizes of posttests for the comparison of notetaking versus control.*



**Figure 2**

*Funnel Plot of studies involving notetaking and control groups*



References

﻿Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing meta-analysis with R: A hands-on guide.* Chapman & Hall/ CRC Press.

Hunter, J.E., Schmidt, F.L., 2004. *Methods of meta-analysis: correcting error and bias in research findings* (2nd ed.). Sage, Thousand Oaks, CA.

Kim, S.K., & Webb, S. (2022). The effects of spaced practice on second language learning: A meta-analysis. *Language Learning, 72*(1), 269–319.

Shintani, N., Li, S., & Ellis, R. (2013). Comprehension-based versus production-based grammar instruction: A meta-analysis of comparative studies. *Language learning, 63*(2), 296­–329.

Viechtbauer, W., Cheung, M.W.-L., 2010. Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1, 112–125. <https://doi.org/10.1002/jrsm.11>

Walters, J., & Bozkurt, N. (2009). The effect of keeping vocabulary notebooks on vocabulary acquisition. *Language Teaching Research, 13*(4), 403–423.

**Interrater Reliability**

**TABLE 1.**Interrater Reliability for Categorical Variables

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Fleiss’s κ | *S* index | Number of coders |
| Publication type | 1 | 1 | 2 |
| L1 | .60 | .63 | 3 |
| L2 | 1 | 1 | 3 |
| Context | 1 | 1 | 2 |
| Region | .87 | .88 | 2 |
| Script | 1 | 1 | 2 |
| Educational level | 1 | 1 | 2 |
| Mode of input | 1 | 1 | 2 |
| Material type | .60 | .60 | 3 |
| Measurement type | 1 | 1 | 2 |
| Learning target | .89 | .91 | 3 |
| Notetaking instruction | .93 | .94 | 3 |
| Instruction type | .94 | .96 | 3 |
| Notetaking behavior | 1 | 1 | 2 |
| Opportunity to review notes | .64 | .74 | 3 |

**TABLE 2.** Interrater Reliability for Continuous Variables

|  |  |  |
| --- | --- | --- |
| Variable | Intraclass correlation | Number of coders |
| Number of notetaking sessions | .99 | 2 |
| Notetaking instruction length | 1 | 2 |
| Test max score | 1 | 2 |
| N | 1 | 2 |
| n | 1 | 2 |
| Post\_mean | 1 | 2 |
| Post\_SD | 1 | 2 |

**References for the Included Studies**

Ahmad, S. Z. (2019). Impact of Cornell notes vs. REAP on EFL secondary school students' critical reading skills. *International Education Studies, 12*(10), 60–74.

Alahmadi, N. S. (2020). The effect of the mind mapping strategy on the L2 vocabulary learning of Saudi learners. *Education and Linguistics Research, 6*(1), 54–68.

Alzu'bi, M. A. (2019). The influence of suggested Cornell note-taking method on improving writing composition skills of Jordanian EFL learners. *Journal of Language Teaching and Research, 10*(4), 863–871.

Aminifard, Y., & Aminifard, A. (2012). Note-taking and listening comprehension of conversations and mini-lectures: any benefit? *Canadian Social Science, 8*(4), 47–51.

Bozorgian, H., & Pillay, H. (2013). Enhancing foreign language learning through listening strategy delivered in L1: An experimental study. *International Journal of Instruction, 6*(1).

Chen, H. J. H., & Yang, T. Y. C. (2013). The impact of adventure video games on foreign language learning and the perceptions of learners. *Interactive learning environments, 21*(2), 129–141.

Hale, G. A., & Courtney, R. (1994). The effects of note-taking on listening comprehension in the Test of English as a Foreign Language. *Language Testing, 11*(1), 29–47.

Hayati, A. M., & Jalilifar, A. (2009). The impact of cultural knowledge on listening comprehension of EFL learners. *English Language Teaching, 2*(1), 101–111.

Jin, Z., & Webb, S. (2021). Does writing words in notes contribute to vocabulary learning? *Language Teaching Research*. Advance online publication.

Kang, E.Y. (2010). Effects of output and note-taking on noticing and interlanguage development. *TESOL and Applied Linguistics, 10*(2). 19­–36.

Kashani, S., & Shafiee, S. (2016). A comparison of vocabulary learning strategies among elementary Iranian EFL learners. *Journal of Language Teaching and Research, 7*(3), 511–518.

Kilickaya, F., & Cokal-Karadas, D. (2009). The effect of note-taking on university students’ listening comprehension of lectures. *Kastamonu Education Journal, 17*(1), 47–56.

Mežek, Š. (2013). Learning terminology from reading texts in English: The effects of note-taking strategies. *Nordic Journal of English Studies, 12*(1), 133–161.

Moradi, S., Ghahari, S., & Abbas Nejad, M. (2020). Learner-vs. expert-constructed outlines: Testing the associations with L2 text comprehension and multiple intelligences. *Studies in Second Language Learning and Teaching, 10*(2), 359­­–384.

Najar, R.L. (1997). *The effects of note taking strategy instruction on comprehension in ESL texts*, (Unpublished doctoral dissertation). Hawaii University.

﻿Piri, A., & Shirkhani, S. (2021). The comparative effects of note-taking and semantic mapping on EFL learners’ vocabulary learning and vocabulary retention. *Journal of new advances in English Language Teaching and Applied Linguistics, 3*(1), 506–517

Uysal, N. M.; Tezel, K.V. (2020). The effects of language learning strategies instruction based on learning styles on reading comprehension. *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi,* (21), 697–714.

Walters, J., & Bozkurt, N. (2009). The effect of keeping vocabulary notebooks on vocabulary acquisition. *Language Teaching Research, 13*(4), 403–423.

Wilberschied, L. (1998). *The relationships among a variation of immediate recall tasks and measures of L1 writing, L2 achievement, and cognitive strategy use in students of high school Spanish* (Unpublished doctoral dissertation). The Ohio State University.

Zarei, A. A., & Adami, S. (2013). The effects of semantic mapping, thematic clustering, and notebook keeping on L2 vocabulary recognition and production. *Journal on English Language Teaching, 3*(2), 17–27.

Zohrabi, M., & Esfandyari, F. (2014). The impact of note taking on the improvement of listening comprehension of Iranian EFL learners. *International Journal of English Language and Literature Studies, 3*(2), 165–175.