

RESEARCH ARTICLE



The effectiveness of note taking through exposure to L2 input: A meta-analysis

Zhouhan Jin and Stuart Webb

University of Western Ontario, London, Ontario, Canada

Corresponding author: Zhouhan Jin; Email: zjin65@uwo.ca

(Received 06 May 2023; Revised 15 September 2023; Accepted 22 September 2023)

Abstract

There has been increasing interest in the effects of note taking in second language (L2) research. However, no meta-analysis has been conducted to examine the relationship between note taking and learning through exposure to L2 input. We retrieved 28 effect sizes from 21 studies ($N = 1992$) to explore the overall effects of note taking as well as to examine the extent to which the effectiveness of note taking is likely to vary as a function of a set of potential moderators (i.e., learner variables, treatment variables, note-taking features, learning target, and measurement type). Results revealed that note taking had a small to medium positive overall effect on learning through exposure to L2 input ($g = 0.56$, 95% CI: 0.24–0.88). Subsequent moderator analyses revealed that variability in the size of note-taking effects across studies was explained by learner variables (context, region, orthographic scripts, institutional level), treatment variables (mode of input, material type), note-taking features (note-taking behavior, number of note-taking sessions, provision and type of note-taking strategy instruction, total length of instruction, opportunity to review notes), learning target, and measurement type. Based on the obtained findings, teachers are recommended to incorporate note taking in L2 classrooms. Pedagogical suggestions and directions for future research are also provided.

Introduction

Note taking has been recognized as the most common learning strategy both in first language (L1; Kobayashi, 2005) and second language (L2; Siegel, 2021) research. To better understand how note taking contributes to learning, researchers have identified two main functions of note taking: encoding and external storage functions (also known as process and product/review effects; Dunkel, 1988). The encoding function occurs as notes are being taken. The physical practice of writing (or typing) notes involves processing input beyond verbatim copying such as perceiving and organizing information as well as relating it to existing knowledge (Rickards & Friedman, 1978). The external storage function, which occurs after the note-taking process is completed, involves reviewing and/or reorganizing and storing the written input into memory to prevent forgetting as well as help retrieve and relearn information that has been

forgotten, serving as the basis for further activities or tests (Armbruster, 2000; Kobayashi, 2006a). Note taking is generally considered to be an effective learning strategy that allows students to encode information and permits later review to stimulate information recall (Dunkel, 1988; Siegel, 2022).

L1 studies have maintained a strong interest in exploring the effectiveness of note taking. Several meta-analyses have been conducted to investigate the effect of taking notes (Kobayashi, 2005), reviewing notes (Kobayashi, 2006a), and note-taking mode (i.e., longhand versus digital note taking; Allen et al., 2020; Voyer et al., 2022) on L1 learning. Furthermore, Morehead, Dunlosky, Rawson, Blasiman, et al. (2019b) conducted a large-scale survey of university students to investigate their note-taking habits and preferences to inform teachers and researchers of how best to support and enhance students' note-taking abilities and address their specific needs.

L2 research investigating the effects of note taking is less common. Several studies have compared one type of note taking (e.g., conventional note taking; Jin & Webb, 2021; Cornell note taking; Hayati & Jalilifar, 2009) with a control condition that did not take notes. Other studies have compared different types of note taking (e.g., conventional note taking versus Cornell note taking; Tsai & Wu, 2010; outline note taking versus conventional note taking; Song, 2012). Studies have also looked at note taking with unimodal input such as reading (Mežek, 2013; Najar, 1997) or listening (Clark et al., 2014; Hale & Courtney, 1994), whereas others looked at note taking through multimodal input such as viewing (Sakurai, 2018). Additionally, several studies have investigated learning through exposure to L2 input to enhance comprehension skills (Hayati & Jalilifar, 2009; Moradi et al., 2020), whereas other studies were directed at L2 learning to improve language ability such as vocabulary knowledge (e.g., Chen & Yang, 2013; Jin & Webb, 2021) and writing skills (Alzu'bi, 2019). The influence of note taking on exposure to L2 input and learning the L2 also led to inconsistent results. For example, taking notes has been found to have a positive effect (Hayati & Jalilifar, 2009), a negative effect (Hale & Courtney, 1994), and no effect (Clark et al., 2014) on listening comprehension. Studies have also found note taking to make both positive contributions (Jin & Webb, 2021) and no contribution to L2 vocabulary learning (Kashani & Shafiee, 2016). Therefore, in L2 research the degree to which note taking contributes to learning as well as the variables such as learning targets that influence the effectiveness of note taking remain unclear.

One way to deepen our understanding of the effects of note taking is to conduct a meta-analysis. Because a relatively small number of L2 studies of note taking have been focused on language learning, this meta-analysis was intended to examine the effects of note taking through exposure to L2 input, including both L2 learning and content learning. Individual studies are restricted, for example, by their research contexts, participant population, and methodological features. In contrast, meta-analysis allows the examination of how the effects of note taking vary in relation to variables. Therefore, this study also investigated the extent to which the effect of note taking is moderated by different variables across studies such as learner characteristics (e.g., institutional level, region), treatment features (e.g., mode of input, material type), note-taking features (e.g., provision of note-taking strategy instruction, opportunity to review notes), learning targets (e.g., linguistic forms, listening), and measurement type (e.g., recognition, recall).

Effects of note taking through exposure to L2 input

A relatively large number of L2 note-taking studies have investigated the effects of conventional note taking (Chen & Yang, 2013; Hale & Courtney, 1994; Jin & Webb, 2021). Conventional note taking is a free-form note-taking behavior in which learners

freely use their desired method to take notes. However, learning gains have been somewhat inconsistent. For example, studies have revealed that for learning L2 vocabulary, the percentages of learning gains were 12.5%–23% (Jin & Webb, 2021) and 31.4% (Chen & Yang, 2013); for listening comprehension, the percentage of learning gains ranged from 32.9% (Wilberschied, 1998) to 72.5% (Hayati & Jalilifar, 2009).

Taking notes after strategy instruction is another highly examined area. Note-taking strategies involve using note-taking formats such as taking outline notes (Moradi et al., 2020; Song, 2012) and Cornell notes (Hayati & Jalilifar, 2009; Tsai & Wu, 2010), using vocabulary notebooks (Zarei & Adami, 2013), and providing learners with teachers' verbal instruction to employ strategies such as attending to the teacher's flow of speech (e.g., pause and tone of voice), organizational cues (e.g., numeric identifiers), and use of abbreviations before students take notes in a treatment (Balaban, 2017; Bozorgian & Pillay, 2013). There is a great deal of variation in the percentage of learning gains in studies investigating learning from taking notes after strategy instruction. For example, Zarei and Adami (2013) found that learners who wrote unfamiliar words in their vocabulary notebooks had gains of 67.8% on a recognition test and of 42.7% on a recall test. Walters and Bozkurt (2009) found that using vocabulary notebooks led to gains of 35.7% on a recall test but gains of only 40.4% on a recognition test. Other types of note-taking instruction that are directed at increasing learners' comprehension through note taking led to differences in the percentage of learning gains. Studies of reading comprehension after note-taking instruction have reported gains of 63.5% (Moradi et al., 2020) and 84.2% (Najar, 1997). There is similar variation in gains in studies of listening comprehension from taking notes after instruction, with results ranging from 44.8% (Aminifard & Aminifard, 2012) to 83.7% (Hayati & Jalilifar, 2009). The variation in the size of gains in both L2 learning and content learning is likely owing to the many differences in learner and treatment variables, note-taking features, learning targets, and measurement types between studies.

Review of moderator variables

Context

Learning settings were categorized into both foreign language (FL) and second language (SL) contexts. Learners in FL contexts tend to have minimal exposure to the target language outside of the classroom, whereas learners in SL contexts tend to have plenty of opportunities to encounter target language both inside and outside of the classroom (Webb & Nation, 2017). No experimental studies have singled out these two settings as an independent variable. Also, studying in FL and SL contexts may be quite different, leading to varying effects of using note taking.

Region

The region in which the study took place might also affect learners' note taking. Siegel and Kusumoto (2022) conducted a survey to compare Japanese and Swedish students' perspectives and habits toward note taking and found cross-cultural differences in note taking. For example, Japanese learners tended to feel that note taking is difficult, whereas fewer Swedish learners reported difficulty in note taking. However, self-reported data can be biased (Wagner, 2015). Aggregating and analyzing data contributed by all empirical research can be more robust.

Orthographic script

Previous L2 research has investigated the effect of orthographic script (Krepel et al., 2021; Zhang & Zhang, 2022). For instance, Spanish and English both have alphabetic scripts, whereas Chinese and English use different scripts. The distance in terms of orthography and representation is much shorter for L1 Spanish than L1 Chinese so it might be easier for Spanish students to learn English because they can depend on their L1 orthographic knowledge as a resource (Pasquarella et al., 2014). A recent meta-analysis also found that sharing the same orthographic script between L1 and L2 allows learners to take advantage of L1 transfer, which may assist both L2 vocabulary acquisition and comprehension (Zhang & Zhang, 2022). Because note taking is likely to involve the transfer of orthographic knowledge, the extent to which learning might be affected by the orthographic script between L1 and L2 would be useful to investigate.

Institutional level

Research has investigated the effects of note taking with participants at different places of study (elementary school, high school, language institute, university level). Kobayashi's (2005) meta-analysis of the encoding effect of note taking on L1 learning revealed that younger learners in primary and secondary schools benefited more from note taking than university students because note taking can compensate for lack of cognitive abilities and skills. However, a recent meta-analysis of listening-strategy instruction on L2 learners' listening performance revealed that older learners tended to benefit more from strategy instruction because they were equipped with superior cognitive and metacognitive abilities (Dalman & Plonsky, 2022). Several meta-analyses examining L2 vocabulary learning also found that older learners tended to make greater learning gains than younger learners (de Vos et al., 2018; Uchiyama et al., 2019). Given that no studies have examined age or place of study as an independent variable, the effects of note taking with L2 learners at different institutional levels remain unclear.

Mode of input

Research has investigated the act of note taking when encountering written input (Najar, 1997), aural input (Hale & Courtney, 1994), and bimodal (Zarei & Adami, 2013) and multimodal input (Chen & Yang, 2013). Different modes of input may affect the encoding function of note taking. For example, students are likely to sequentially alternate between reading and note taking, whereas note taking during listening involves only one step of simultaneous processing (Kiewra, 1991). This might lead to higher learning gains from reading than from listening because reading provides opportunities for learners to adjust their reading rate to improve their concentration and organize their notes when they skim and reread specific types of information (Slotte & Lonka, 1999). In contrast, note taking during listening involves multiple cognitive processes (Armbruster, 2000; Piolat et al., 2005). Students must pay attention to a lecture, temporarily capture the important information provided by the instructor, hold and organize these ideas in working memory, and simultaneously write them down before they are forgotten.

Material type

Note-taking and L2 studies consist of both academic (Hayati & Jalilifar, 2009; Zarai & Adami, 2013) and nonacademic input (Chen & Yang, 2013; Jin & Webb, 2021). Academic input such as academic lectures, discussions, conversations, and dialogues, the topics of which are academic in nature, may be more complex and dense than nonacademic input such as everyday conversations and teacher's anecdotes (e.g., personal stories and experiences) that might be considered both informative and entertaining (Uysal & Tezel, 2020). Thus, academic input may require students to engage in higher level thinking and deeper cognitive processing than nonacademic input (Tsai & Wu, 2010). Therefore, the complexity of academic input may pose a challenge for note taking. Because little attention was given to the effects of note taking between these two types of input, it is useful to examine material type as a variable to determine whether it affects learning.

Note-taking session

Note-taking studies involve participants taking notes within a single session or across multiple sessions. For example, Najar (1997) allowed students to practice taking notes over nine sessions and found positive encoding effects. In contrast, Hale and Courtney (1994) did not find a positive encoding effect within a single session; however, Jin and Webb (2021) found that taking notes within a single session positively affected vocabulary learning. Examining number of sessions as a moderator may reveal the degree to which it influences the effects of note taking.

Note-taking behavior

Research investigating the use of note taking in the L2 context is typically categorized into two different scenarios: required note taking, which stipulates that students must take notes, and voluntary note taking, which allows students to take notes at their own discretion. Voluntary note taking may represent the most typical learning situation because students take notes as they choose (Hale & Courtney, 1994). Students who voluntarily choose to take notes are likely to be motivated and engaged, which may contribute to positive learning outcomes (Koren, 1997). However, students who are required to take notes may benefit less from the encoding function of note taking because they may be less motivated and engaged with the note-taking process. Hale and Courtney (1994) conducted the only study comparing these two types of note-taking behavior. They found that voluntary note taking had little effect on listening comprehension and required note taking negatively affected their learning performance. However, the negative effect of note taking does not mean there is less value of considering this variable, because many L2 primary studies have revealed a positive effect of voluntary note taking (Jin & Webb, 2021; Wilberschied, 1998) and required note taking (Hayati & Jalilifar, 2009; Kang, 2010). A meta-analysis of studies with learners who have completed voluntary note taking and required note taking can shed more light on this variable.

Provision of note-taking strategy instruction

Even though most L2 learners acknowledge the value of note taking (Armbruster, 2000), they often produce poor notes (e.g., verbatim transcript), which decreases the learning benefits of note taking. One way to make up for shortcomings in students'

personal notes is to employ note-taking strategies (Siegel, 2022) such as taking notes using a specific framework (e.g., Cornell notes) or recording key words and any useful linguistic forms. However, the results of note-taking-strategy instruction have been inconsistent. For example, after providing note-taking-strategy instruction, Kang (2010) found a large positive note-taking effect ($g = 1.17$) on L2 vocabulary learning. Similarly, Najjar (1997) found a large positive note-taking effect ($g = 1.11$) on reading comprehension. However, Aminifard and Aminifard (2012) found little difference between note taking after instruction and no note taking on listening comprehension, whereas Kashani and Shafiee (2016) found a negative effect of note taking after instruction on L2 vocabulary learning.

Instruction time

As noted by Plonsky and Oswald (2014), examining the correlation between treatment length and outcomes would be useful for considering the cost/benefit ratio of interventions. Several meta-analyses have investigated length of strategy instruction (SI) in domains such as pragmatics instruction (Plonsky & Zhuang, 2019) and listening SI (Dalman & Plonsky, 2022). Therefore, the synthetic approach of this study enables us to examine the relationship between the effectiveness of note taking and the overall duration of instruction time.

Opportunity to review notes

In L1 research, participants who had the opportunity to review their notes produced higher learning gains than those who did not have the opportunity to review their notes (Kiewra, 1991; Kobayashi, 2006a). Several studies of L2 note taking have included the opportunity for participants to review notes (Hayati & Jalilifar, 2009; Najjar, 1997; Walters and Bozkurt, 2009), whereas other studies have not included review of notes (Jin & Webb, 2021; Kashani & Shafiee, 2016; Moradi et al., 2020). However, no L2 studies have explicitly investigated the effects of reviewing notes on learning. Thus, the degree to which reviewing notes influences learning through exposure to L2 input remains unclear.

Learning target

Research has investigated the effects of note taking on L2 listening comprehension (Hayati & Jalilifar, 2009; Tsai & Wu, 2010), reading comprehension (Moradi et al., 2020; Najjar, 1997), vocabulary learning (Jin & Webb, 2021; Zarei & Adami, 2013), grammatical structures (Kang, 2010), and writing skills (Alzu'bi, 2019). The degree to which note taking may affect different learning targets has yet to be examined.

Measurement type

The effect of note taking may vary depending on the measurement. L2 vocabulary research (Laufer & Goldstein, 2004) indicated that learners tend to score higher on recognition tests (e.g., multiple-choice items) than recall tests (e.g., write the meaning of the given word or write the L2 word that corresponds with a given meaning). Similarly, Walters and Bozkurt (2009) found that note takers learned more words on recognition tests than on recall tests. However, in L1 research, Kobayashi's (2005) meta-analysis

examining the encoding effects of note taking found that the benefits of note taking were greater when learning was measured by recall tests than by recognition tests. This finding is consistent with those of an earlier L1 study by Weener (1974). Thus, there is value in synthesizing prior note-taking studies to determine the influence of measurement type in the L2 context.

Research questions

This study has two main purposes. One is to better understand the overall effect of note taking on learning through exposure to FL or SL input, and the second is to explain potential moderators of the note-taking effect. To this end, two research questions (RQs) were formulated:

1. What is the overall effect of note taking through exposure to L2 input?
2. To what extent does the effectiveness of note taking vary across learner variables, treatment variables, note-taking features, learning target, and measurement type?

Method

Literature search

Following guidelines on literature search for meta-analyses (In'nami & Koizumi, 2010; Plonsky, 2015), we comprehensively searched the following databases: ProQuest (including data subsets such as PsycINFO, ProQuest Dissertations and Theses, Linguistics and Language Behavior Abstracts, Education Resources Information Center), MLA International Bibliography, ScienceDirect, Social Sciences Citation Index. Additionally, Google and Google Scholar were searched. The following keywords were used to search for studies: (a) tak* notes, notetaking, note-taking, (b) learning, comprehension, vocabulary, grammar, and (c) second language, foreign language, and L2. Further, we manually reviewed journals that are widely cited in second language acquisition (SLA) and applied linguistics including *Applied Linguistics*, *Canadian Modern Language Review*, *Computer Assisted Language Learning*, *Foreign Language Annals*, *Language Learning*, *Language Learning and Technology*, *Language Teaching*, *Language Teaching Research*, *Modern Language Journal*, *Reading in a Foreign Language*, *Second Language Research*, *Studies in Second Language Acquisition*, *System*, and *TESOL Quarterly*. In addition, the reference sections in reviewed studies were carefully examined, resulting in the addition of two more studies (Clark et al., 2014; Sakurai, 2018). We set July 2022 as the completion point for data collection. A PRISMA flow diagram (Page et al., 2021) outlining the study selection process can be found in the [Supplementary Materials](#).

Selection criteria

The following criteria were applied:

1. The input must be a second or foreign language to the participants.
2. One of the independent variables measured must be note taking. This may include one, or a combination, of the following forms of note taking: taking notes using one's own method, using a note-taking format such as Cornell notes, or taking notes after instructional interventions, meaning that learners were instructed in a combination of note-taking strategies such as how to identify main ideas and use abbreviations.

3. At least one dependent variable measured some aspect of L2 or content learning (e.g., listening comprehension, vocabulary learning).
4. When measuring vocabulary learning, the study must have focused on learning single-word items, not multiword items because only one study (Jin & Webb, 2021) examined multiword items.
5. The studies must be experimental or quasi-experimental and include a control condition in which participants received the same treatment but did not take notes. A true control group that only completed tests was not considered in our study because none of the primary studies included such a group.
6. Only studies adopting between-participants designs were included. Studies that used within-participants designs were excluded due to a limited number of studies using within-participants designs ($k = 7$; Carrell et al., 2004; Clark et al., 2004). This decision was made based on the suggestion that combining both within- and between-participants design in meta-analyses is likely to produce biased results (Plonsky & Oswald, 2015).
7. Studies had to control for participants' preexisting knowledge. Control could occur through the use of pretests revealing no statistically significant difference between groups on pretest scores (e.g., listening comprehension skill: Bozorgian & Pillay, 2013; vocabulary knowledge: Jin & Webb, 2021; reading comprehension skill: Najar, 1997); or if the pretest was not administered, the participants were randomly assigned (Kang, 2010); or their proficiency level had been checked and found to be homogenous (Kashani & Shafiee, 2016).
8. The studies must have been written in English.
9. The studies reported enough statistical information for an effect size to be calculated (i.e., mean, *SD*, and the number of participants tested).
10. The full text of the article was available.
11. The studies did not focus on learners with language learning problems.

We contacted the authors and gratefully received additional information from Uysal and Tezel (2020). After applying the inclusion criteria to the retrieved reports, a total of 21 studies ($N = 1992$) met the criteria and were included in the analysis. The studies consisted of 19 published journal articles and two doctoral dissertations. As only three studies reported delayed posttest results (Jin & Webb, 2021; Kashani & Shafiee, 2016; Piri & Shirkhani, 2021), our data focused exclusively on the results of immediate posttests.

Coding

All studies were coded as specified by the coding scheme table. The table illustrates 27 variables, including 14 moderator variables. Coded features were divided into seven main categories pertaining to study identification, learner variables, treatment variables, note-taking features, outcome features, methodology, and study quality (see [Supplementary Materials](#)).

To assess the reliability of our coding procedure, 10 studies (approximately 47.6% of 21 studies) were randomly selected and independently coded by three raters.¹ All three

¹The subset of 10 studies was first recoded by a second rater. Additionally, based on suggestions from two reviewers, several new moderators were added to the study. Consequently, both the newly added moderators and some previously coded moderators were assessed by a third coder. For more detailed information, please refer to the [Supplementary Materials](#).

raters are experts in one area of second language acquisition and had also carried out meta-analyses. The average agreement using Fleiss's Kappa between three coders was .90. The S index (Norouzian, 2021) was calculated at the item level of categorical moderators, and intraclass correlation was used for the continuous variables (see [Supplementary Materials](#) for item-level estimates). All discrepancies were discussed and resolved.

Learner variables

Context was coded as FL and SL. Region was also coded. At first it was coded as countries such as China, Iran, Sweden, and USA. Due to a limited number of studies in each country, we combined countries according to larger regions (i.e., Asia, Middle East, North America, Europe). The orthographic script used in learners' L1 and L2 was coded as (a) same, (b) different, or (c) mixed. Institutional level was coded as elementary school, secondary school, language institute, and university.

Treatment variables

Mode of input consisted of three categories: reading, listening, and mixed. The label *mixed* included treatments in which bimodal or multimodal input was included. Material type was coded as either (a) academic or (b) nonacademic.

Note-taking features

The variable note-taking behavior was coded based on whether students voluntarily took notes or were required to take notes. The second note-taking feature was provision of note-taking-strategy instruction. We coded this variable as *absence* for studies that did not provide any note-taking strategies. That is, students used conventional note taking, which is taking notes using their desired method. Conversely, we coded the variable as *presence* when one or more note-taking strategies were provided. The provided note-taking strategy was further divided into linear learning strategy, generative learning strategy, and unreported. Although previous research (Bui & Myerson, 2014; Kobayashi, 2005; Siegel, 2022) commonly categorized note taking into "verbatim" and "generative" methods, this study employed the terms "linear" and "generative" to describe two distinct types of note-taking strategies. Following Ponce and Mayer (2014), a linear learning strategy involves processing information in the same structure presented in material, whereas a generative learning strategy involves selecting and reorganizing information to build a coherent structure. That said, a linear learning strategy is more focused on selecting and recording information following the flow of the material that is encountered without necessarily attempting to copy written or spoken input word for word and a generative learning strategy allows for more flexibility in how the information is organized. For example, using vocabulary notebooks and encouraging learners to write down key vocabulary or any useful linguistic forms was categorized as a linear learning strategy. Strategies such as encouraging learners to use abbreviations and record the gist of the material, break down long sentences into shorter sentences, or develop a specific note-taking format (e.g., Cornell notes) were categorized as generative learning strategies. Although outline notes are also called outline linear (Chen et al., 2017), constructing outline notes in this study was considered as a generative learning strategy because learners were allowed to personalize their outlines by using their own words, abbreviations, and symbols (Moradi et al., 2020).

Opportunity to review notes

This variable was initially coded as review, no review, and unreported. However, we acknowledge that the quality of students' notes may affect the usefulness of reviewing (Kobayashi, 2006a) and the provision of strategy instruction on note taking might enhance their note quality. To account for this, we further subdivided the studies into two groups based on the quality of notes: conventional note taking and note taking after strategy instruction. This allowed us to investigate the potential differences in the effect of reviewing notes between students who received instruction and those who did not. By doing so, we sought to provide a more nuanced understanding of how this variable might moderate note-taking effects on learning in the L2 context.

Outcome variables

For dependent variables, maximum score for the test, mean posttest scores, *SDs*, total number of participants, and participant number in each group were coded. The learning target was coded as linguistic forms, listening, reading, or writing. Learning linguistic forms referred to vocabulary learning except one study (Kang, 2010), which examined the combination of grammar and vocabulary learning. Listening and reading consisted of listening and reading comprehension, respectively. One study (Alzu'bi, 2019) looked at the influence of note taking on improving learners' writing skills.

Measurement type

The type of measurement was initially categorized into different types of instruments such as meaning recall, form recall, multiple-choice tests, and True/False questions. Following two meta-analyses in L1 note taking (Kobayashi, 2005; Voyer et al., 2022), recall (e.g., meaning recall, form recall) and recognition (e.g., multiple-choice test, True/False questions) were coded as two distinct measures. Moreover, one study (Alzu'bi, 2019) investigated students' improvement of writing skills, which is different from recall or recognition of information; therefore, the measurement type of this study was coded as free writing. Also, the results of Zohrabi and Esfandyari's (2014) study did not differentiate the types of measurement, so the label *mixed* was included. To summarize, the measurement type was finally divided into categories of recall, recognition, writing, mixed, and unreported.

Number of note-taking sessions and length of note-taking instruction

These two moderators were continuous variables. It is important to note that note-taking sessions and instruction time might not cover the whole treatment sessions (e.g., Bozorgian & Pillay, 2013). Therefore, only treatment related to note taking was considered and the total length of instruction time was coded in minutes.

Calculation of effect size

The first step of calculating the effect size was to use standardized mean difference, which is called Cohen's *d*. Then, we converted Cohen's *d* to the unbiased effect size Hedges's *g* by multiplying a correction factor: $J = 1 - [3(4 \times df - 1)]$. The reason for selecting Hedges's *g* (Hedges & Olkin, 1985) as the effect size is because it is an unbiased estimate of effect sizes compared with Cohen's *d*, especially when samples are small (Borenstein et al., 2009). To interpret the values for effect size for this study, we referred

to Plonsky and Oswald's (2014) benchmarks for defining the magnitude of effect sizes: small = 0.40, medium = 0.70, and large = 1.00. The correlation coefficient was used to examine the role of the two continuously measured moderators (length of note-taking instruction, number of note-taking sessions), and the magnitude of effect sizes was small = .25, medium = .40, and large = .60.

To ensure independence of effects and minimize the presence of sample size inflation, all effect sizes pertaining to the same participants were averaged to form a single effect size for each sample. However, for studies reporting data on one or more treatment groups and a control condition (Hayati & Jalilifar, 2009), effect sizes were calculated by contrasting each treatment group with the control condition on the immediate posttest (see Norris & Ortega, 2000, p.446). To answer RQ2, if a study reported multiple measurement types, these effect sizes were kept separate in analyzing the moderating effect of the type of outcome measures on the relationship in question.

Data analysis

We conducted all the analyses using the meta package (Schwarzer, 2007) in the R statistical environment (R Core Team, 2020). To address RQ1, we aggregated the effect sizes from the studies that compared note taking with no note taking (i.e., control condition) to produce a weighted mean effect size. When doing so, the effect sizes were weighted by inverse variance so that those with less sampling error contributed more to the meta-analytic mean (Plonsky & Oswald, 2015). To answer RQ2, we conducted subgroup analyses for moderator variables to determine whether the coded learner characteristics, treatment variables, note-taking features, L2 learning target, and measurement type (which all served as categorical moderator variables) as reported by the data set were significant moderators of the effectiveness of note taking. Considering the possibility that our categorization of subgroups might introduce new sampling errors at the subgroup level, random-effects modeling was used for between-subgroup comparisons while controlling for such sampling errors (Harrer et al., 2019; Plonsky & Oswald, 2015; Suzuki et al., 2021) and a between-group *Q*-statistic was used to examine the influence of moderator variables on effect sizes (ESs). For the two continuous moderators (number of note-taking sessions, length of note-taking instruction), the corresponding data were subjected to meta-regression analyses. Additionally, following Plonsky and Zhuang (2019), a correlational approach was adopted to examine their relationship with note-taking effects.

Sensitivity analyses

After aggregating the effect sizes, we detected outliers to ensure the robustness of the results and assessed publication bias that might influence the current data sets. It is important to mention that studies identified as potential outliers do not necessarily mean the study is an outlier that does not reflect normal language learning because each study was independently conducted and included a different group of students and varying learning conditions (Yanagisawa & Webb, 2021). Therefore, we followed Viechtbauer and Cheung's (2010) guidance and reran the whole analysis while excluding the studies identified as potential outliers and compared the results with those that were obtained when including all studies (see [Supplementary Materials](#) for publication bias and sensitivity analyses results).

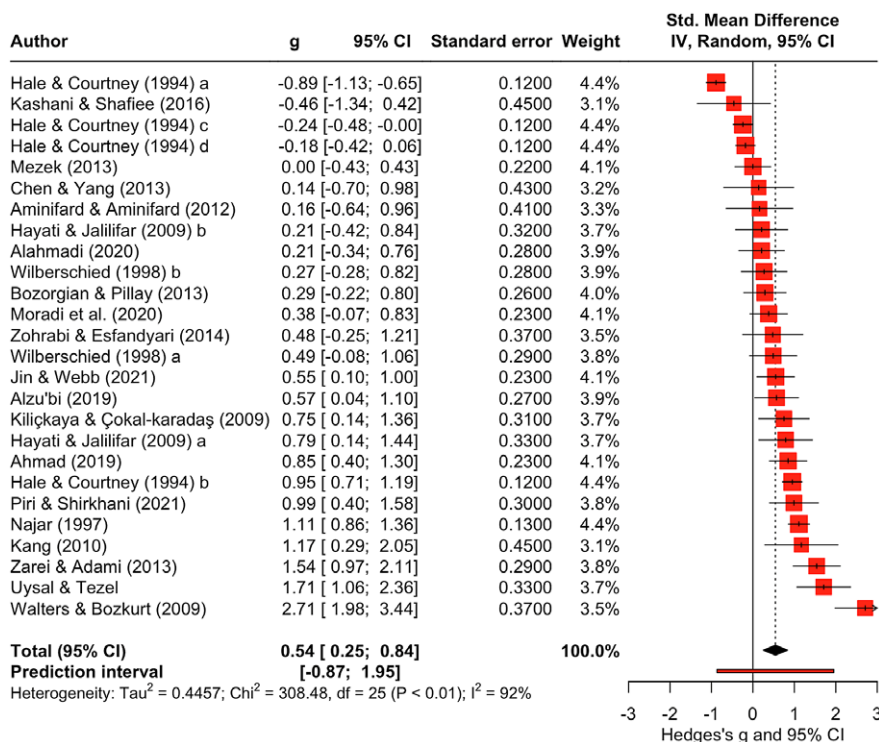


Figure 1. Forest plot for the posttest results of the comparison of note taking versus control.

Results

RQ1: What is the overall effect of note taking through exposure to L2 input?

To answer the first RQ, 26 aggregated effect sizes were included (Figure 1). The results showed that there was significantly greater learning for note-taking conditions than for non-note-taking control conditions. However, the homogeneity test was statistically significant, $Q(24) = 308.94$, $p < .001$, indicating that variability in the true effect across studies as well as sampling error could have created this difference. The overall mean effect size from the posttest results for note taking versus no note taking was 0.56 ($p = .002$, 95% CI $[0.24, 0.88]$), a small to medium effect size according to Plonsky and Oswald's benchmarks (2014; between-group contrast, 0.4 for small, 0.7 for medium, and 1.0 for large). We then conducted moderator analyses to examine the extent to which different factors could account for this variability (see Table 1 for a detailed description of the subgrouped effect sizes in each category).

RQ2: What variables moderate the effect of note taking?

Learner variables

Four factors relating to participant characteristics were examined: (a) context, (b) region, (c) orthographic script, and (d) institutional level. The analyses showed that

Table 1. Moderator analyses (categorical variables)

Variables	<i>k</i>	<i>n</i>	ES (Hedges's <i>g</i>)	95% CI		<i>Q</i> _{between} (<i>p</i>)
				Lower	Upper	<i>Q</i> _{between} (<i>p</i>)
Learner variables						
<i>Context</i>						
FL	18	19	0.69	0.36	1.03	2.69 (.101)
SL	3	7	0.18	−0.48	0.84	
<i>Region</i>						
Asia	3	3	0.70	−0.47	1.87	3.68 (.298)
Middle East	11	12	0.54	0.24	0.85	
North America	3	7	0.18	−0.48	0.84	
Europe	4	4	1.27	−0.60	3.14	
<i>Orthographic scripts</i>						
Same	5	6	0.96	−0.10	2.03	2.32 (.314)
Different	14	15	0.58	0.32	0.84	
Mixed	2	5	0.12	−0.94	1.18	
<i>Institutional level</i>						
University	11	15	0.34	0.02	0.66	8.92 (.030)
Language institute	7	7	1.13	0.28	1.97	
Secondary	2	3	0.57	−0.18	1.32	
Elementary	1	1	−0.46	−1.34	0.42	
Treatment variables						
<i>Mode of input</i>						
Listening	9	14	0.27	−0.03	0.57	5.84(.054)
Reading	7	7	0.85	0.34	1.37	
Mixed	5	5	0.84	−0.74	2.43	
<i>Material type</i>						
Academic	15	19	0.51	0.13	0.89	0.21 (.645)
Nonacademic	6	7	0.64	0.12	1.15	
NT features						
<i>NT behavior</i>						
Allowed	7	9	0.29	−0.15	0.73	2.69 (.260)
Required	14	16	0.71	0.28	1.14	
Unreported	1	1	0.29	−0.22	0.80	
<i>Provision of NT strategy instruction</i>						
Presence ^a	15	15	0.84	0.43	1.25	7.20 (.007)
Absence ^b	7	11	0.16	−0.20	0.53	
<i>Type of NT strategy</i>						
Linear learning strategy	5	5	1.04	−0.47	2.56	0.27 (.875)
Generative learning strategy	8	8	0.75	0.35	1.15	
Unreported	2	2	0.78	−2.41	3.97	
<i>Opportunity to review notes</i>						
Yes	13	17	0.55	0.14	0.96	2.04 (.361)
No	6	6	0.35	−0.12	0.81	
Unreported	3	3	0.94	−0.77	2.65	
<i>Conventional vs. control</i>						
Review	4	7	0.10	−0.49	0.69	0.62 (.734)
No review	3	3	0.33	−0.45	1.10	
Unreported	1	1	0.14	−0.70	0.98	
<i>NT after strategy instruction vs. control</i>						
Review	9	10	0.88	0.33	1.43	3.15 (.207)
No review	3	3	0.37	−1.56	2.29	
Unreported	2	2	1.27	−2.22	4.76	

(Continued)

Table 1. (Continued)

Variables	<i>k</i>	<i>n</i>	ES (Hedges's <i>g</i>)	95% CI		<i>Q</i> _{between} (<i>p</i>)
				Lower	Upper	<i>Q</i> _{between} (<i>p</i>)
Learning target						7.22 (.065)
Linguistic forms	9	9	0.76	0.03	1.49	
Listening	7	12	0.22	-0.12	0.57	
Reading	4	4	0.99	0.14	1.83	
Writing	1	1	0.57	0.04	1.10	
Measurement type						2.49 (.647)
Recognition	11	16	0.50	0.05	0.95	
Recall	7	8	1.07	0.27	1.88	
Free writing	1	1	0.57	0.04	1.10	
Mixed	1	1	0.48	-0.25	1.21	
Unreported	2	2	0.46	-2.96	3.89	

Note. *k* = number of studies; *n* = number of effect sizes; NT = note taking.

^aAny provided note-taking strategies directed at improving students' note-taking performance such as introducing a specific note-taking format.

^bConventional note taking, meaning that students take notes using their own method without receiving any note-taking strategies.

Table 2. Descriptive statistics of continuous moderators

	Mean	SD	Min	Max	<i>k</i>
Number of NT sessions	4.60	5.07	1	20	21
Total amount of NT instruction time (minutes)	109.08	165.26	0	510	16

Note. *k* refers to the number of primary studies (among a total of 21 studies) that reported the corresponding information; NT = note taking.

the mean effect size for studies conducted in FL contexts ($g = 0.69$) was larger than that for studies in SL contexts ($g = 0.18$) and note-taking effect was the largest in Europe ($g = 1.27$), followed by Asia ($g = 0.70$), the Middle East ($g = 0.54$), and North America ($g = 0.18$). In terms of orthographic script, it was found that learners whose L1 has the same script as the L2 yielded a higher effect size in note taking ($g = 0.96$) than those whose L1 and L2 scripts were different ($g = 0.58$). Also, the note-taking effect was found to be the lowest in a diverse group of learners including those with the same script as the L2 and those without ($g = 0.12$). In relation to participants' institutional level, the effect size of note taking from language institute students ($g = 1.13$) was the largest, followed by those for secondary school ($g = 0.57$) and for university students ($g = 0.33$). Also, note taking among elementary school students negatively affected L2 learning ($g = -0.09$), indicating that note taking might hinder learning for young learners. However, the effect size of secondary school and elementary school students should be interpreted with caution due to the small sample of each.

Treatment variables

Regarding mode of input, the differences in effect size between note taking through mixed input ($g = 0.84$) and reading ($g = 0.85$) were small. Both input modes tended to lead to greater learning than did taking notes when listening ($g = 0.27$). In addition, studies using nonacademic materials produced a larger effect than studies using academic materials ($g = 0.64$ and $g = 0.51$, respectively).

Note-taking features

The results showed that requiring students to take notes tended to produce higher learning gains ($g = 0.71$) than allowing them to take notes ($g = 0.29$). The effect of note taking is substantially larger when note-taking instruction was involved ($g = 0.84$) than without note-taking instruction ($g = 0.16$). Moreover, the effectiveness of note taking on learning varied depending on the type of note-taking strategy. A linear learning strategy was more effective ($g = 1.04$) than a generative learning strategy ($g = 0.75$).

In addition, the opportunity to review notes yielded a higher effect size ($g = 0.55$) than no review of notes ($g = 0.35$). It should be noted that after the use of note-taking strategies, there was a substantial difference between those who reviewed ($g = 0.88$) and those who did not review notes ($g = 0.33$). However, when participants had the opportunity to review notes without note-taking instruction, the effect size was smaller ($g = 0.10$) than when participants did not review notes ($g = 0.33$).

Learning target

Medium to large effects were obtained for treatment groups over control conditions in reading ($g = 0.99$) and learning linguistic forms ($g = 0.73$), a small to medium effect was found for writing ($g = 0.57$), and a more modest effect was found for listening ($g = 0.22$).

Measurement type

The largest effect was found with recall tests ($g = 1.07$). Somewhat smaller positive effects were observed among studies investigating other types of measurement such as writing a composition ($g = 0.57$) and taking recognition tests ($g = 0.50$).

Number of note-taking sessions and note-taking instruction length

Number of note-taking sessions and total amount of instruction time were two continuous variables (see Table 2 for descriptive statistics). Preserving the continuous nature of these two moderators is intended to build a more precise model of the correlation between the cost-ratio benefit of note taking in class and the effects of note taking. The results showed a medium to large positive correlation between number of note-taking sessions and note-taking effects ($r = 0.51$). For total amount of instruction time (in minutes), however, the correlation was small ($r = 0.21$). Meta-regression analyses also revealed that number of note-taking sessions was a significant predictor of note-taking effects ($p < .001$) but total amount of instruction time was not ($p = .336$).

DISCUSSION

The analysis of the overall effects of note taking revealed that learning through note taking was significantly more effective than learning without note taking ($g = 0.56$). This finding is important because it clarifies the value of note taking with respect to learning through exposure to L2 input. Several meta-analyses have already been conducted concerning the effects of note taking on L1 learning (Kobayashi, 2005, 2006a, 2006b), but a relatively small number of intervention studies from L2 contexts have focused on the effects of note taking (Siegel, 2022). However, the effect size found in this meta-analysis was larger than those that were reported in the earlier L1 meta-analyses. Kobayashi (2005), for example, included 57 studies and found a mean weighted effect size of 0.22 for note taking versus no note taking. Therefore, the larger influence of note

taking on learning through exposure to L2 input found in the present study reveals an area that warrants further attention.

However, it should, also be noted that the effect of note taking in this study was on both language learning and content learning. Previous meta-analyses that investigated the effects of other types of strategies were focused solely on L2 learning such as corrective feedback ($d = 0.64$; Li, 2010) and spacing ($g = 0.58$; Kim & Webb, 2022), and these strategies were found to be slightly more effective than note taking. One possible explanation for the modest effect of note taking could be the absence of instruction from teachers on developing learners' note-taking skills. Almost half of the included studies did not involve instruction in any note-taking strategies. However, this may have ecological validity because Siegel (2019) reported that a considerable proportion of teachers may spend little time teaching note-taking skills. If students take poor notes, they may not effectively encode information and so may not fully benefit from taking and reviewing notes (Kobayashi, 2005).

The homogeneity assumption was found to be violated, $Q(21) = 280.53$, $p < .001$, indicating that the observed differences in effect sizes across studies could be due to both variability in true effects and sampling error. Therefore, potential moderating variables were further analyzed, and the effect of note taking was found to vary across a range of moderators pertaining to learner, treatment, note-taking features, L2 learning targets, and measurement type.

Moderator analysis

In relation to learner variables, the analyses revealed greater benefit from note taking in FL contexts ($g = 0.68$) over SL contexts ($g = 0.12$). This pattern of results aligns with findings from other meta-analyses that explored different SIs such as listening SI (Dalman & Plonsky, 2022) and reading SI (Taylor, 2014). A possible reason for this is that for FL learners taking notes may be more effective for improving learning (Siegel, 2018b) because note taking provides opportunities to optimize learning from the small amount of L2 input that FL learners typically encounter. However, because SL learners encounter large amounts of input inside and outside of the classroom, the effect of note taking might be mediated by many other positive variables such as frequency of encounters, which is a primary determinant of language acquisition and processing (Ellis, 2002).

The results also showed that learners in Europe ($g = 1.27$) tended to benefit more from note taking than those in Asia ($g = 0.70$), the Middle East ($g = 0.54$), and North America ($g = 0.18$). For FL learners, there was a substantial difference in note-taking effects obtained in Europe relative to Asia and the Middle East. One possible reason may be due to cross-cultural differences between perceptions of note taking. Siegel and Kusumoto (2022) reported a cross-cultural investigation of note taking between Sweden and Japan. The study investigated note taking from various perspectives such as L2 note takers' views about note taking and education systems. For example, Japanese students tended to feel that note taking was a challenging activity, whereas Swedish students did not. The cross-cultural differences might also be attributed to orthographic transparency between L1 and L2. It is possible for L2 learners to record information in their L1 and/or L2 (Siegel, 2021). As information about language use in notes was limited to the included studies, exploring translanguaging in student notes was beyond the scope of this study. However, the results for the orthographic script moderator variable provide some indication that sharing the same orthographic script ($g = 0.96$) provided greater learning benefit than did sharing different orthographic scripts ($g = 0.58$). This finding is

consistent with Zhang and Zhang's (2022) meta-analysis indicating that sharing the same orthographic script could contribute better to learning through exposure to L2 input. The magnitude of effect sizes in this study is larger than the effect sizes from Zhang and Zhang's (2022) study, perhaps due to additional benefits of note taking.

With respect to the significantly moderating effect of institutional level, the results showed a substantial difference in the effects of note taking from language institutes ($g = 1.13$) relative to secondary ($g = 0.57$) and postsecondary schools ($g = 0.34$). This finding is in line with a previous meta-analysis that examined listening SI on L2 listening comprehension (Dalman & Plonsky, 2022). This result may relate to higher motivation to learn a target language for language institute students who choose to study an L2 compared with secondary school or university students who may be required to study an L2. Interestingly, it was found that note taking negatively affected learning for elementary school students ($g = -0.46$). Note taking is a complex and demanding activity (Piolat et al., 2005), which might present a significant challenge for younger learners. However, because there was only one sample included at the level of elementary school students, further research is warranted to test this interpretation.

Of the treatment moderator variables, the analyses revealed that learning gains were larger for note taking with mixed ($g = 0.84$) and written input ($g = 0.85$) than with aural ($g = 0.27$) input. Smaller effects of note taking on listening might be expected because it is difficult for learners to adjust their focus and organize their notes due to time pressure (Bui & Myerson, 2014). The analyses also revealed that the effectiveness of note taking is more pronounced when using nonacademic input ($g = 0.64$) than when using academic input ($g = 0.51$), suggesting that students can derive greater benefits from their note-taking efforts by incorporating nonacademic materials into their learning process. Although academic materials are often perceived as the primary source for learning through exposure to L2 input, our results indicated that nonacademic materials also play a crucial role in enhancing the learning experience through note taking.

Regarding note-taking features, a medium effect size was found when requiring students to take notes ($g = 0.71$), whereas a small effect size was found when allowing students to take notes ($g = 0.29$). This can be expected because students who were required to take notes were more likely to pay attention and engage with the material being presented. However, the findings contrast with the results that were obtained by Hale and Courtney (1994) who found little effect of voluntary note taking and a negative effect of required note taking on listening comprehension. Further research investigating how these two note-taking behaviors affect learning would be useful.

The results also revealed a significant positive effect of instruction on note-taking strategies ($g = 0.84$), suggesting that learners may not know how to effectively take notes. Instruction in note-taking strategies can help students more effectively select, organize, and elaborate on information in their notes (Siegel, 2018a). This is supported by Mayer's (1996) SOI (i.e., selection, organization, integration) model of learning, which suggests that meaningful learning occurs when learners are able to select relevant information, organize it into a structure in working memory, and integrate it with prior knowledge from long-term memory. When looking at different types of note-taking strategies, the results showed that a linear learning strategy ($g = 1.04$) led to greater learning than a generative learning strategy ($g = 0.75$). This finding might be surprising given the general optimism that surrounds the effectiveness of learning from generative note taking (e.g., Siegel, 2022). The larger effect for linear learning strategies is likely due to the focus on learning in this meta-analysis. Learners who used a linear learning strategy may direct more attention to unfamiliar linguistic forms, whereas in generative learning strategies there is additional focus on making decisions about the importance of content

and deciding when and how to take notes (Siegel, 2021). This is supported by input processing theory, which states that learners' processing of input is largely determined by what they attend to (Han & Peverly, 2007). It should also be noted that both types of note-taking instruction are helpful for enhancing learners' ability to process input, as the effect sizes for note taking are medium and large in generative and linear learning strategy, respectively.

The results also suggested that reviewing notes may not yield positive results unless learners were provided with note-taking-strategy instruction. After note-taking instruction, students who reviewed notes showed greater learning gains ($g = 0.88$) than those who did not have opportunities to review notes ($g = 0.37$). Therefore, providing note-taking instruction and allowing students to review their notes could optimize L2 learning. As no L2 empirical research to date has explicitly examined the effect of students reviewing their own notes after they have completed the process of note taking, this finding may provide some insights into the role of the external storage function of note taking in the L2 context.

In relation to different learning targets, note-taking effects appeared to be larger for reading ($g = 0.99$) and linguistic forms ($g = 0.76$), followed by writing ($g = 0.57$) and listening ($g = 0.22$). This pattern of results is similar to those of Plonsky (2011) that the effectiveness of SI is larger in reading ($g = 0.74$) and vocabulary ($g = 0.64$) than in writing ($g = 0.42$) and listening ($g = 0.06$). Therefore, this finding builds on previous research by demonstrating the significance of note taking in enhancing specific language skills, particularly in reading comprehension and linguistic forms. However, because there is only one study that investigated note-taking and writing skills, further research in this area is warranted.

For the moderator of measurement type, the analysis found that note-taking effects were more pronounced when learning was measured by recall tests ($g = 1.07$) than by writing ($g = 0.57$) and recognition tests ($g = 0.50$). This is consistent with an earlier meta-analysis of L1 note taking that indicated that the benefits of note taking were greater when measured with recall tests than with recognition tests (Kobayashi, 2005). This suggests that note taking can be an efficient strategy for helping learners to successfully recall newly learned knowledge from memory because recall tests tend to be a relatively demanding test format in comparison with recognition test formats (Laufer & Goldstein, 2004). The depth-of-processing hypothesis provides support for the effectiveness of note taking in that greater depth of processing results in superior recall (Craik & Tulving, 1975) and note taking is found to bring about deeper processing of information (Titsworth & Kiewra, 2004). Although writing a composition typically involves higher level processing, investigating note-taking effects when measured by writing skills in future research is crucial due to the limited number of studies ($k = 1$).

Regarding the length of instruction, the results revealed a small but positive correlation ($r = 0.21$) between duration of instruction and note-taking effects, suggesting that longer instruction is more effective than shorter instruction. However, the small correlation indicated that the benefits of extensive instruction time may be limited. Additionally, a medium to large positive correlation was found between number of note-taking sessions and note-taking effects ($r = 0.51$), suggesting that the frequency or consistency of note-taking practice might be more important than the length of instruction time. This is supported by skill acquisition theory (e.g., DeKeyser, 2007), which suggests that repeated and focused practice as well as revision of skills is important to develop abilities such as note taking. That said, with limited class time, integrating note-taking practice as an ongoing component of instruction and providing regular opportunities for students to engage with this skill may yield favorable learning outcomes.

It is also important to note that the results of the present study focused on the magnitude of the observed difference instead of the direction (i.e., p values) to provide a more nuanced understanding of the results. Within moderator analyses, comparisons are often underpowered, so the findings need to be interpreted with caution.

Limitations and future directions

This meta-analysis provided preliminary evidence that note taking has a small to medium positive effect on learning through exposure to L2 input. Given that note taking is not confined to L2 classrooms, the findings of this study may shed light on the broader implications of the transferability of note-taking skills across languages, disciplines, and beyond educational context (i.e., work context).

However, several limitations were identified that would be useful to address in future research. First, it would be useful for additional studies to include (1) younger learners such as elementary and secondary school students, (2) comparisons of different note-taking types such as outline versus Cornell note taking, (3) greater focus on other L2 learning targets such as grammar learning, and (4) the use of digital note taking on laptops or tablets. It should be noted that this meta-analysis specifically focused on longhand note taking because no studies to date have compared learning in digital note taking versus a control condition. It is also worth noting that digital note taking has received significant attention in L1 research (e.g., Luo et al., 2018; Morehead, Dunlosky, & Rawson, 2019a; Mueller & Oppenheimer, 2014) and should also be given appropriate attention in L2 research.

Moreover, researchers mainly looked at note taking through reading or listening. However, with the development of technological advancements, the use of multimodal input such as viewing for L2 learning has become increasingly common (Feng & Webb, 2020; Webb, 2015). Only one study to date (Sakurai, 2018) has investigated L2 learners' note taking through viewing.

Also, we would like to encourage researchers to consider reporting information recorded in student notes. It would be useful for teachers and researchers to better understand the role of translanguaging in note taking in the L2 context. The bilingual spelling in alphabetic systems (BAST) model (Tainturier, 2019) proposed that the level of coactivation in bilingual spelling is influenced by the extent of similarity between the orthographic and phonological aspects of the two languages. Moreover, the relative proportion of high and low similarity in word form and meaning between two languages can affect the degree of facilitation or inhibitory effects that occur when encoding information (Inieta et al., 2021). Therefore, future research is recommended to investigate the influence of cross-linguistic similarity on note taking and learning through exposure to L2 input.

Regarding study quality in L2 research, out of 21 studies, 15 (71%) reported pretest scores, 8 (38%) reported instrument reliability, and only 3 (14%) included a delayed posttest. Plonsky and Zhuang (2019) found that higher study quality is correlated with larger effect sizes. There is a need for future research to increase and report methodological quality (e.g., instrument reliability, large sample size, pretest and delayed posttest inclusion, random assignment) to make studies stronger and yield larger effects.

Finally, it is recommended that a minimum of three studies is included in each moderator subgroup (Li, 2016; Vuogan & Li, 2023). However, Plonsky (2011) suggested that the prospective value of a meta-analysis sometimes might be greater than its retrospective value. Therefore, given the limited number of L2 studies that have

examined the effects of note taking on learning, we did not exclude categories with less than three ESs. Although these analyses are meaningful because they indicate the trends in the data, further research on note-taking effects is needed to reach firmer conclusions.

Supplementary material. The supplementary material for this article can be found at <http://doi.org/10.1017/S0272263123000529>.

Data availability statement. The experiment in this article earned Open Data and Open Materials badges for transparent practices. The materials and data are available at <http://doi.org/10.48316/p9RfC-1IHfQ>

Acknowledgments. We would like to express our sincere gratitude to the Handling Editor Luke Plonsky and three anonymous reviewers for their constructive feedback that helped improve our manuscript significantly. We are also grateful to Lizz Huntley and Andrew Dennis for their great support during our manuscript submissions. We also thank the following researchers for providing assistance and feedback for the current meta-analysis: Akifumi Yanagisawa, Takumi Uchihara, Emi Iwaizumi, Yanxue Feng, and Dr Michael Sutton.

References

- Allen, M., LeFebvre, L., LeFebvre, L., & Bourhis, J. (2020). Is the pencil mightier than the keyboard? A meta-analysis comparing the method of notetaking outcomes. *Southern Communication Journal*, 85, 143–154.
- Aminifard, Y., & Aminifard, A. (2012). Note-taking and listening comprehension of conversations and mini-lectures: Any benefit? *Canadian Social Science*, 8, 47–51.
- Armbruster, B. B. (2000). Taking notes from lectures. In R. F. Flippo & D. C. Caverly (Eds.), *Handbook of college reading and study strategy research* (pp. 175–199). Lawrence Erlbaum Associates.
- 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, 863–871.
- Balaban, S. (2017). *A comparison of shadowing and note-taking as two while-listening strategies and their effects on EFL learners* [Unpublished master's thesis]. Yeditepe University, İstanbul, Turkey.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. (2009). *Introduction to meta-analysis*. Wiley.
- Bozorgian, H., & Pillay, H. (2013). Enhancing foreign language learning through listening strategies delivered in L1: An experimental study. *International Journal of Instruction*, 6, 105–122.
- Bui, D. C., & Myerson, J. (2014). The role of working memory abilities in lecture note-taking. *Learning and Individual Differences*, 33, 12–22.
- Carrell, P. L., Dunkel, P. A., & Mollaun, P. (2004). The effects of note-taking, lecture length, and topic on a computer-based test of ESL listening comprehension. *Applied Language Learning*, 14, 83–105.
- Chen, P. H., Teo, T., & Zhou, M. (2017). Effects of guided notes on enhancing college students' lecture note-taking quality and learning performance. *Current Psychology*, 36, 719–732.
- 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, 129–141.
- Clark, M., Wayland, S., Osthus, P., Brown, K. G., Castle, S., & Ralph, A. (2014). *The effects of note taking on foreign language listening comprehension*. University of Maryland Center for Advanced Study of Language. <http://www.govtilr.org/Publications/Notetaking.pdf>
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104, 268–294.
- Dalman, M., & Plonsky, L. (2022). The effectiveness of second-language listening strategy instruction: A meta-analysis. *Language Teaching Research*. <https://doi.org/10.1177/13621688211072981>
- DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 97–113). Lawrence Erlbaum.
- de Vos, J. F., Schriefers, H., Nivard, M. G., & Lemhöfer, K. (2018). A meta-analysis and meta-regression of incidental second language word learning from spoken input. *Language Learning*, 68, 906–941.
- Dunkel, P. (1988). The content of L1 and L2 students' lecture notes and its relation to test performance. *TESOL Quarterly*, 22, 259–281.

- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143–188.
- Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening, and viewing: Which mode of input is most effective? *Studies in Second Language Acquisition*, 42, 499–523.
- 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, 29–47.
- Han, Z., & Peverly, S. T. (2007). Input processing: A study of ab initio learners with multilingual backgrounds. *International Journal of Multilingualism*, 4, 17–37.
- Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2019). *Doing meta-analysis with R: A hands-on guide* (1st ed.). Chapman and Hall/CRC.
- Hayati, A. M., & Jalilifar, A. (2009). The impact of cultural knowledge on listening comprehension of EFL learners. *English Language Teaching*, 2, 101–111.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.
- Iniesta, A., Rossi, E., Bajo, M. T., & Paolieri, D. (2021). The influence of cross-linguistic similarity and language background on writing to dictation. *Frontiers in Psychology*, 12, Article 679956.
- In'nami, Y., & Koizumi, R. (2010). Database selection guidelines for meta-analysis in applied linguistics. *TESOL Quarterly*, 44, 169–184.
- Jin, Z., & Webb, S. (2021). Does writing words in notes contribute to vocabulary learning? *Language Teaching Research*. Epub ahead of print 14 December 2021.
- Kang, E. Y. (2010). Effects of output and note-taking on noticing and interlanguage development. *TESOL and Applied Linguistics*, 10, 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, 511–518.
- Kiewra, K. A. (1991). Aids to lecture learning. *Educational Psychologist*, 26, 37–53.
- Kim, S. K., & Webb, S. (2022). The effects of spaced practice on second language learning: A meta-analysis. *Language Learning*, 72, 269–319.
- Kobayashi, K. (2005). What limits the encoding effect of note-taking? A meta-analytic examination. *Contemporary Educational Psychology*, 30, 242–262.
- Kobayashi, K. (2006a). Combined effects of note-taking/-reviewing on learning and the enhancement through interventions: A meta-analytic review. *Educational Psychology*, 26, 459–477.
- Kobayashi, K. (2006b). Conditional effects of interventions in notetaking procedures on learning: A meta-analysis. *Japanese Psychological Research*, 48, 109–114.
- Koren, S. (1997). Listening to lectures in L2: Taking notes in L1. *TESL-EJ*, 2, 1–16.
- Krepel, A., de Bree, E. H., & de Jong, P. F. (2021). Does the availability of orthography support L2 word learning? *Reading and Writing*, 34, 467–496.
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54, 399–436.
- Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, 60, 309–365.
- Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition*, 38, 801–842.
- Luo, L., Kiewra, K. A., Flanigan, A. E., & Peteranetz, M. S. (2018). Laptop versus longhand note taking: Effects on lecture notes and achievement. *Instructional Science*, 46, 947–971.
- Mayer, R. E. (1996). Learning strategies for making sense out of expository text: The SOI model for guiding three cognitive processes in knowledge construction. *Educational Psychology Review*, 8, 357–371.
- Mežek, Š. (2013). Learning terminology from reading texts in English: The effects of note-taking strategies. *Nordic Journal of English Studies*, 12, 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, 359–384.
- Morehead, K., Dunlosky, J., & Rawson, K. (2019a). How much mightier is the pen than the keyboard for note-taking? A replication and extension of Mueller and Oppenheimer (2014). *Educational Psychology Review*, 31, 753–780.
- Morehead, K., Dunlosky, J., Rawson, K., Blasiman, R., & Hollis, R. (2019b). Note-taking habits of 21st century college students: Implications for student learning, memory, and achievement. *Memory*, 27, 807–819.

- Mueller, P. A., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. *Psychological Science*, 25, 1159–1168.
- Najar, R. L. (1997). *The effects of note taking strategy instruction on comprehension in ESL texts* [Unpublished doctoral dissertation]. Hawaii University.
- Norouzian, R. (2021). Interrater reliability in second language meta-analyses: The case of categorical moderators. *Studies in Second Language Acquisition*, 43, 896–915.
- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50, 417–528.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, 10, 1–11.
- Pasquarella, A., Deacon, H., Chen, X., Commissaire, E., & Au-Yeung, K. (2014). Acquiring orthographic processing through word reading: Evidence from children learning to read French and English. *International Journal of Disability, Development and Education*, 61, 240–257.
- Piolat, A., Olive, T., & Kellogg, R.T. (2005). Cognitive effort during note taking. *Applied Cognitive Psychology*, 19, 291–312.
- 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, 506–517.
- Plonsky, L. (2011). The effectiveness of second language strategy instruction: A meta-analysis. *Language Learning*, 61, 993–1038.
- Plonsky, L. (Ed.). (2015). *Advancing quantitative methods in second language research*. Routledge.
- Plonsky, L., & Oswald, F. L. (2014). How big is big? Interpreting effect sizes in L2 research. *Language Learning*, 64, 878–912.
- Plonsky, L., & Oswald, F. L. (2015). Meta-analyzing second language research. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 106–128). Routledge.
- Plonsky, L., & Zhuang, J. (2019). A meta-analysis of second language pragmatics instruction. In N. Taguchi (Ed.), *Routledge handbook of SLA and pragmatics* (pp. 287–307). Routledge.
- Ponce, H. R., & Mayer, R. E. (2014). Qualitatively different cognitive processing during online reading primed by different study activities. *Computers in Human Behavior*, 30, 121–130.
- R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Rickards, J. P., & Friedman, F. (1978). The encoding versus the external storage hypothesis in note taking. *Contemporary Educational Psychology*, 3, 136–143.
- Sakurai, S. (2018). Promoting skills and strategies of lecture listening and note-taking in L2. *Fukuoka University Journal of Humanities*, 49, 1019–1046.
- Schwarzer, G. (2007). meta: An R package for meta-analysis. *RNews*, 7, 40–45.
- Siegel, J. (2018a). Did you take “good” notes? On methods for evaluating student notetaking performance. *Journal of English for Academic Purposes*, 35, 85–92.
- Siegel, J. (2018b). Teaching lecture notetaking with authentic materials. *ELT Journal*, 73, 124–133.
- Siegel, J. (2019). Notetaking in English language teaching: Highlighting contrasts. *TESOL Journal*, 10, 1–5.
- Siegel, J. (2021). *Developing notetaking skills in a second language: Insights from classroom research*. Routledge.
- Siegel, J. (2022). Research into practice: Teaching notetaking to L2 students. *Language Teaching*, 55, 245–259.
- Siegel, & Kusumoto, Y. (2022). A cross-cultural investigation of L2 notetaking: student habits and perspectives. *Journal of Multilingual and Multicultural Development*. Advance Online Publication, 1–17.
- Slotte, V., & Lonka, K. (1999). Review and process effects of spontaneous note-taking on text comprehension. *Contemporary Educational Psychology*, 24, 1–20.
- Song, M. Y. (2012). Note-taking quality and performance on an L2 academic listening test. *Language Testing*, 29, 67–89.
- Suzuki, S., Kormos, J., & Uchihara, T. (2021). The relationship between utterance and perceived fluency: A meta-analysis of correlational studies. *The Modern Language Journal*, 105(2), 435–463.
- Tainturier, M. J. (2019). A theory of bilingual spelling in alphabetic systems. In T. Olive & C. Perret (Eds.), *Spelling and writing words: Theoretical and methodological advances* (pp. 67–92). Brill.

- Taylor, A. M. (2014). L1 glossing and strategy training for improving L2 reading comprehension: A meta-analysis. *International Journal of Quantitative Research in Education*, 2, 39–68.
- Titsworth, B. S., & Kiewra, K. A. (2004). Spoken organizational lecture cues and student notetaking as facilitators of student learning. *Contemporary Educational Psychology*, 29, 447–461.
- Tsai, T.-F., & Wu, Y. (2010). Effects of note-taking instruction and note-taking languages on college EFL students' listening comprehension. *New Horizons in Education*, 58, 120–132.
- Uchiyama, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, 69, 559–599.
- 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.
- Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1, 112–125.
- Voyer, D., Ronis, S. T., & Byers, N. (2022). The effect of notetaking method on academic performance: A systematic review and meta-analysis. *Contemporary Educational Psychology*, 68, Article 102025.
- Vuogan, A., & Li, S. (2023). A systematic review of meta-analyses in second language research: Current practices, issues, and recommendations. *Applied Linguistics Review*. Advance online publication.
- Wagner, E. (2015). Survey research. In B. Paltridge, & A. Phakiti (Eds.), *Research methods in applied linguistics* (pp. 83–100). Bloomsbury.
- Walters, J., & Bozkurt, N. (2009). The effect of keeping vocabulary notebooks on vocabulary acquisition. *Language Teaching Research*, 13, 403–423.
- Webb, S. (2015). Extensive viewing: Language learning through watching television. In D. Nunan & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 159–168). Routledge.
- Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.
- Weener, P. (1974). Notetaking and student verbalization as instructional learning activities. *Instructional Science*, 3, 51–74.
- 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.
- Yanagisawa, & Webb, S. (2021). To what extent does the involvement load hypothesis predict incidental L2 vocabulary learning? A meta-analysis. *Language Learning*, 71, 487–536.
- 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, 17–27.
- Zhang, S., & Zhang, X. (2022). The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis. *Language Teaching Research*, 26, 696–725.
- 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, 165–175.