8-2022

The effect of typewriting vs. handwriting lecture notes on learning: a systematic review and meta-analysis.

Timothy Schaun Lau
University of Louisville

THE EFFECT OF TYPEWRITING VS HANDWRITING LECTURE NOTES ON LEARNING: A SYSTEMATIC REVIEW AND META-ANALYSIS

By

Timothy Schaun Lau
B.A., University of Nevada Las Vegas, 2009
M.S., Johns Hopkins University, 2011

A Dissertation
Submitted to the Faculty of the
College of Education and Human Development of the University of Louisville
In Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy
in Counceling and Personnel Services

Department of Counseling and Human Development
University of Louisville
Louisville, Kentucky

August 2022

THE EFFECT OF TYPEWRITING VS HANDWRITING LECTURE NOTES ON LEARNING: A SYSTEMATIC REVIEW AND META-ANALYSIS

By

Timothy Schaun Lau
B.A., University of Nevada Las Vegas, 2009
M.S., Johns Hopkins University, 2011

A Dissertation Approved on

29 June 2022

by the following Dissertation Committee:

____
Dissertation Director
Jeff Valentine

____
Marci DeCaro

____
Jason Immekus

____
Patrick Pössel

ACKNOWLEDGMENTS

Sincerely, earnestly, and humbly, thank you Jeff Valentine for your assistance and investment of time and energy in me and my ideas as I slowly progressed towards this achievement. Of all the nice things I could say about you, I think what's been most impressive to me about you is not just all your substantial accomplishments, but how you did it, lovingly, kindly, patiently. You didn’t only teach me how to do high quality research, you showed me how to be a better man.

I also thank Dr. Marci DeCaro, Dr. Jason Immekus, and Dr. Patrick Pössel for their support throughout my doctoral career and on this dissertation specifically.

ABSTRACT

THE EFFECT OF TYPEWRITING VS. HANDWRITING LECTURE NOTES ON LEARNING: A SYSTEMATIC REVIEW AND META-ANALYSIS

Timothy Schaun Lau

29 June 2022

This study is a systematic review and meta-analysis of studies examining the effect of note-taking modality during lecture, that is, taking notes by hand using pen and paper vs. taking notes using a keyboard and computer, on learning among secondary and postsecondary students. I begin with a review of the literature and theoretical introduction to the theories and terms used. From a theoretical standpoint, there are strong reasons to believe that taking notes by hand might offer recall benefits relative to taking notes using a computer and keyboard. At the same time, I point out that one problem, which I term the "fundamental problem of modality research", is that when researchers randomly assign participants to a note-taking modality they are also, indirectly, assigning them to a note-taking style. Furthermore, most studies do not consider factors such as participant transcription capacity that might serve as theoretically important moderators.

I then describe the methods used for the systematic review and meta-analysis. These included a robust literature search, double screening of all potentially eligible studies, and double coding of all eligible studies. The meta-analytic methods involved multilevel applications of standard meta-analytic methods. The systematic review

resulted in identification of 33 eligible reports containing 42 independent samples and 88 effect sizes, all evaluating whether there are recall differences — almost always operationalized as scores on a quiz given after exposure to lecture material — between participants taking notes by handwriting vs. typewriting, that is, the modality effect. A statistically significant overall meta-analytic average was found g = +0.144 [0.023, 0.265], p = .021, benefiting handwriters over typewriters. This is a small effect; on average, in the typical study typewriters scored about 50% on the recall quiz. The effect size of g = +0.14 translates into an average percent correct of about 57% in the handwriting group. There is some evidence that providing participants with an opportunity to review their notes might substantially reduce the observed advantage for handwriters.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

OVERVIEW OF STUDY

This study is a systematic review and meta-analysis of studies examining the effect of note-taking modality, that is, taking notes by hand using pen and paper vs. taking notes using a keyboard and computer, on learning among secondary and postsecondary students. I begin with a review of the literature and theoretical introduction to the theories and terms used. From a theoretical standpoint, there are strong reasons to believe that taking notes by hand (i.e., handwriting) might offer recall benefits relative to taking notes using a computer and keyboard (i.e., typewriting). At the same time, I point out that one problem, which I term the "fundamental problem of modality research," is that when researchers randomly assign participants to a note-taking modality they are also, indirectly, assigning them to a note-taking style. Furthermore, most studies do not consider factors such as participant transcription capacity that might serve as theoretically important moderators.

I then describe the methods used for the systematic review and meta-analysis. These included a robust literature search, double screening of all potentially eligible studies, and double coding of all eligible studies. The meta-analytic methods involved multilevel applications of standard meta-analytic methods. The systematic review resulted in identification of 33 eligible reports containing 42 independent samples and 88 effect sizes, all evaluating whether there are recall differences — almost always operationalized as scores on a quiz given after exposure to lecture material — between participants taking notes by handwriting versus typewriting, that is, the modality effect. A statistically significant overall meta-analytic average was found $g = +0.144$ [0.023, 0.265], $p = .021$, benefiting handwriters over typewriters. This is a small effect; on average, in the typical study typewriters scored about 50% on the recall quiz. The effect size of $g = +0.14$ translates into an average percent correct of about 55.9% in the handwriting group. There is some evidence that providing participants with an opportunity to review their notes might substantially reduce the observed advantage for handwriters.

I then discuss three limitations and consequent directions for future research arising from my systematic review. These are: the extent to which the studies included in this review are practically relevant, the extent to which the studies included in this review are theoretically relevant, and the fact that there was a non-trivial degree of effect size heterogeneity that was largely unexplained. I conclude by offering recommendations to note-takers based on the results of this systematic review.

CHAPTER II

INTRODUCTION

Note-taking is a catchall term for a complex system of processes — involving, for example, visual and/or aural perception and comprehension, graphomotor movement, and meta-cognition (Flavell, 1979) — that govern how information is transcribed into a format that can be accessed later, all with the intent of reducing information loss. Note-taking is one of the earliest cognitive technologies (Dror, 2007) and is a ubiquitous human activity. Some examples of note-taking are courtroom stenographers, who use a specialized tool and vocabulary in an attempt to capture verbatim court proceedings; reporters, who might use a pen and notebook to write down interview responses; and students, who handwrite or type notes during a lecture to facilitate learning. In classroom contexts, note-taking in response to lecture formatted learning can be traced back as far as the German Protestant universities of the 16th century (Clark, 2008). As an example of the ubiquity of note-taking in classroom settings, Landrum, 2010, estimated that 80% of undergraduate class time is devoted to lectures, during which students are expected to take notes.

Technological advancements in the medium through which note-taking is performed (e.g., chisel, pencil, digital stylus) have progressively optimized the editability, storability, portability, and accessibility of notes, as well as the rapidity with which those notes can be created. Advancements associated with personal computers created a new storage device for note-taking (i.e., electronic file format) and simultaneously have created several new note-taking modalities (e.g., brain, Li et al., 2017; gaze, Špakov & Miniotas, 2004 and Hansen et al., 2004; gesture, Manikandan et al., 2018; speech, and keyboard to digital text vs. handwriting on paper). The transition from notes written by hand, using pen and paper, to typewritten notes with keyboard and computer has been the subject of hundreds of comparison research studies. Apart from the practical benefits associated with understanding the effects of note-taking modality on recall, there has been debate among theorists regarding the benefits of handwritten vs. typed notes. Briefly, some theorists (e.g., Mueller & Oppenheimer, 2014) contend that handwritten notes improve learning because this modality lends itself to a summative note-taking style, that is, notes that reflect the note-taker's synthesis of the information. Typists are likely to attempt to take approximately verbatim notes (Muller & Oppenheimer, 2014; see experiment two), which by definition involves less self-generated content. The extra processing resulting from self generated content while transcribing notes should lead to improved encoding and subsequent recall. Other theorists argue that verbatim notes are more effective because post-note-taking they contain more information and are therefore better study and review aids (Bui, Myerson, & Hale, 2013).

Apart from these theoretical considerations, there might be practical reasons for one modality to be superior to another, depending on context. For example, when transcribing symbols and formulas typewriting might be significantly more cumbersome than handwriting. Similarly, if the transcribed materials would benefit from spatial strategies (e.g., concept maps), handwriting is likely to be substantially better than typing (Fiorella & Mayer, 2017).

Empirical studies, as well as replication studies, examining the effects of note-taking modality on recall (e.g., Morehead, Dunlosky, & Rawson, 2019; Schoen, 2012; Gür, 2021) and even meta-analyses (see Allen et al., 2020 & Voyer et al., 2022) have produced mixed results. Allen concluded with a statistically significant averaged r = -.142 in favor of handwriting over "electronic devices" and gave explanations and pedagogical recommendations in light of the findings. While Voyer et al. (2022) found a, not statistically significant, mean estimated g = -0.008, concluding that "...an apparent advantage of longhand notetaking reported in some previous studies can be explained at least partially by distractions from notetaking by other applications that are present only with digital devices." Both reviews included studies with varying devices (e.g., soft keyboards & stylus') and input from varying sources (e.g., text, audio, video) creating additional heterogeneity between what were already incomplete set's of studies. As a result, a state-of-the-art systematic review and meta-analysis has yet to be done on this problem. I expect my results will differ from theirs both because of the constraints of study level factors and the robust literature search.

Literature Review

Note-taking is an act of distributed cognition (Hollan, Hutchins, & Kirsh, 2000) also described as transactive memory (Wegner, 1986) viz., the distribution of memories, computation, or information represented on an accessible external artifact. Note-taking involves attending to, perceiving, decoding, encoding, and temporarily storing what is then transcribed. In essence, note-taking is an activity designed to ease the cognitive burden of information processing, sometimes known as cognitive offloading (Risko & Gilbert, 2016; Boldt & Gilbert, 2019). Short-term memory capacity is limited, so any information that is not stored in long-term memory will be displaced when new information is attended to. But during a lecture, for example, students have no hope of being able to attend to, perceive, decode, encode, temporarily store, and then store in long-term memory all the relevant information. Note-taking solves this problem for students by offloading part of this burden. The notes can then be used to help get more of the information into long term memory later through review. The benefit of creating an external store to help with later long-term memory storage has a cost: Piolat, Olive, and Kellog, (2004) point out that note-taking splits attention between transcription and listening and/or watching. They also suggest that attention switching from aural and visual stimuli to note-taking places additional cognitive load on the note-taker, and therefore might detract from encoding and subsequent recall. At the same time, provided sufficient working memory given the pace at which the information is received, note-taking can provide additional opportunities to encode and possibly synthesize the information in real time, both of which should lead to better recall.

As implied by the foregoing discussion, note-taking does not exist in a bubble: the note-taking process is a responsive system between the note-taker and their environment. When students take notes during lecture, for example, they are responding to both external and internal influences by recording information to help them meet their goals (e.g., to perform well on an exam). Internal influences include the student's transcription capacity, information-processing capacity, working memory capacity, attention, primary language, familiarity or expertise with the observed information, intended purpose for taking notes, and past experience with and preferences for particular note-taking methods and tools. External influences include the information source (e.g., whether it requires attention to multiple stimuli); note-taking modality (e.g., typing and writing); whether or not the material can be reviewed (asynchronously later, or synchronous review[1] when the notes are being taken); whether learning aids are present and effective (e.g., skeletal or guided notes and instructor notes); the pace, intermittence, and complexity of the incoming information; as well as constraints imposed by the chosen note-taking style (i.e., summative and verbatim). Table 1 identifies and describes some of these internal and external influences — the most important of these are also described in detail in the "*Potential Moderators of the Note-taking Modality Effect*" section below. It is notable that in most studies of the note-taking process, all these dimensions could potentially vary across participants and conditions, but usually, only a few are explicitly acknowledged and controlled. Many of the internal influences listed can be addressed in experiments that use randomization to assign participants to conditions but because these influences are often not well articulated, the external influences will likely vary in known and unknown ways across studies and therefore will likely complicate between study comparisons.

Furthermore, study characteristics are sometimes confounded with each other, which makes it harder to determine why a particular empirical result was observed. For example, when participants are assigned to the typing modality they will tend to take verbatim notes (Mueller & Oppenheimer, 2014; see experiment 2). As previously mentioned, a verbatim note-taking style probably influences encoding and the resulting recall measurements. It is then unclear whether any differences observed in recall are due to differences in note-taking style, differences in processing associated with note-taking

[1] This is a difficult factor to control for and (as far as I can tell) has never been controlled for in a note-taking study.

style, or their interaction. I therefore refer to this type of confounding of factors in a dynamical system as the fundamental problem of modality research.

Research on Note-taking Modality

Research on note-taking can be traced back to the early 1900's. Initially, this research focused on the effectiveness of note-taking (H. E. Jones, 1923; Crawford, 1925; E. S. Jones, 1927; Greene, 1928; 1934; Wagner & Strabel, 1935; McClendon, 1956; Vernon, 1946; Ash & Carlton 1951; Eisner & Rohde, 1959). Note-taking was often considered a potential distraction, splitting the note-taker's attention between listening and transcription. Consensus that note-taking can improve recall eventually emerged in the 1970's (DiVesta & Gray, 1972; Hartley & Davies, 1978), culminating with a meta-analysis by Kobiyashi (2006). When general consensus on this topic emerged, note-taking efficacy research branched into several different sub-topics. Importantly, note-taking modality has been an active and growing topic of research as the personal computer continues to become the standard writing medium in educational settings.

More recently, a modality study by Mueller and Oppenheimer (2014) generated a flurry of interest in the popular media. Examples of the study's impact in the popular media include the *Washington Post* ("Handwriters learn better, hands down"; Barbash, 2014), *The Atlantic* ("Students do worse on quizzes when they use keyboards in class"; Meyer, 2014), and *Scientific American* ("Don't Take Notes with a Laptop"; May, 2014) and was followed by numerous educator decisions prohibiting the use of computers in the classroom (e.g., Gross, 2014; Strauss, 2016). In a series of experiments using college students, Mueller and Oppenheimer (2014) observed the effects of note-taking modality (i.e., typing vs. writing) on factual and conceptual recall while controlling for note-taking style and review. In their first experiment, typists and writers performed equally well (d = 0.07) on factual questions but on conceptual questions writers had statistically significantly better recall than typists (d = 0.20). In the second experiment a condition was added in which typists were encouraged not to take verbatim notes. The authors observed that the encouragement intervention was "completely ineffective" (p. 5)[2] at changing typists' proclivity to take verbatim notes and subsequently did not make a difference in the conceptual recall advantage found in experiment one. In the third experiment, a condition was added in which all note-takers, regardless of modality, were given the opportunity to review their notes. Across three outcome types (i.e., factual, conceptual, and combined) writers performed better than typists (d = 0.15), though this finding was not statistically significant. However, for all three outcome types there was a statistically significant modality by review interaction, such that handwriters who had the opportunity to review their notes had better recall than participants in the other three conditions (d's = 0.20 - 0.40). The authors suggested that the improved recall in the writing condition could be attributed to information being synthesized or processed more deeply by the act of summative note-taking.

In addition to generating interest in the popular media, Mueller and Oppenheimer (2014) has since been the subject of several (not entirely successful) replication studies (e.g., Kirkland, 2016; Luo et al., 2018; Mitchell & Zheng, 2017; Morehead et al., 2019). For example, Morehead et al. replicated and extended the methods of Mueller and Oppenheimer but failed to find consistent evidence supporting the original findings.

[2] Even with the intervention being ineffective, one might suppose that if note-taking style were a moderating factor, aligning the writing styles would reduce the observed effect of modality in outcomes. The data shows that it did (see corrigendum), even with poor adherence to treatment protocol.

These contradictory findings highlight the need for a state-of-the-art systematic review and meta-analysis.

Theoretically Important Potential Moderators of the Note-taking Modality Effect

Note-taking modality is one of many potential moderators affecting the relationship between note-taking and recall. Below, I focus on six moderators that are based on cognitive theory: the note-taker's transcription speed, the speaker's pace, the note-taker's note-taking style, whether there is an opportunity to review notes, the note-takers' intentions, and how recall was measured. I focus on these moderators because they are theoretically important in the sense that they can have important effects on the extent to which note-taking accomplishes the goal of reducing information loss. With the exception of the note-taker's transcription speed (which is sometimes used as a control variable within studies), all of these moderators can vary across studies and, if sufficiently reported, were used in the analyses reported below.

Transcription Capacity. When taking notes in an academic setting from a lecture, the amount of time available to process information is a product of the interaction between speaker pace (information going in) and transcription speed (information going out). Regarding transcription speed, typical handwritten transcription speed (Peverly, 2006) varies by gender, age, primary language, socio-economic status, and education (Van Drempt, McCluskey, & Lannin, 2011; O'Mahony et al., 2008; Piolat, Barbier, & Roussey, 2008). Words per minute (WPM) is the colloquial metric for evaluating the flow of words; standardized at 5 letters in English (Arif & Stuerzlinger, 2009), including spaces and punctuation. Typical adults' handwriting speed ranges between 5-25 WPM (Bledsoe, 2011; Connelly, Dockrell, & Barnett, 2005; Brown, 1988; Rogers & Case-

Smith, 2002; Karat et al., 1999) while typical typewritten transcription ranges between 37-77 WPM (Karat et al., 1999; Dhakal et al., 2018). The speed differential between these two modalities is to be expected, because handwriting is a unimanual (i.e., one-handed) process, whereas typing is bimanual (i.e., two-handed), both doubling manual transcription input capacity and minimizing the graphomotor movement necessary to produce letters. Proficient typists are much faster than proficient writers, but that speed may come with a cognitive cost (Bouriga & Olive, 2021), specifically, less resources for processing the incoming information.

Another factor related to transcription capacity relevant to creating digital notes is the distinction between hardware and software keyboards. Hardware keyboards are switching stations that use the mechanical movement of the keys (i.e., travel; the distance a key moves down when pressed) to provide haptic feedback to users. Software keyboards are virtual representations of keyboards that can — but do not universally — provide vibrations as haptic or sounds as audio feedback. There are known differences in transcription capacity between hardware and software keyboards. Due to their less consistent and robust haptic feedback, software keyboards are associated with slower transcription speed, higher error rates, and the use of additional cognitive resources (Arif & Stuerzlinger, 2009; Paek, 2008; Mackenzie & Soukoreff, 2002; Mackenzie & Zhang, 1999; Soukoreff & Mackenzie, 1995), though of course as the technology evolves and typists become more comfortable with this particular entry medium, the differences between hardware and software keyboards may diminish.

Speaker Pace. Speaker pace, or more generally, the pace at which information is delivered to students, is a variable that has the potential to affect the quality of notes and hence, student learning (Robinson et al., 1997). Neither handwriters nor typewriters with average transcription capacity can keep up with the average university lecture pace: the typical spoken rate for university lecture depends on a number of factors such as rhythmic patterning, stress, language, and pauses, but ranges from 100-200 WPM (Bain, Basson, & Wald, 2002). Most students taking handwritten notes have no hope of creating an accurate verbatim record. The differential between typical handwriting speed and typical rate of speech during lectures may affect student meta-cognitive decisions regarding note-taking style. That is, speaker pace interacts with transcription capacity in the sense that if transcription capacity is high and speaker pace low, the note-taker can choose between different note-taking styles (summative and verbatim). If transcription capacity is low and speaker pace is high, the note-taker can only choose between taking summative notes and fragmentary verbatim notes.

Note-taking Style. Note-taking style refers to whether notes are intended to be a more or less complete transcription of the information (i.e., verbatim notes) or whether they are intended to represent the notetaker's synthesis of information (i.e., summative notes). The primary feature of summative notes is that they reflect meta-cognitive decisions regarding what information should be transcribed. Creating summative notes therefore requires note-takers to self-generate more content than verbatim notes, whereas verbatim notes are more transcriptive in nature. Therefore, summative notes require cognitive resources to be focused on comprehension and synthesis. Verbatim notes enable the note-taker to stop synthesizing the incoming information and focus on transcribing an approximately accurate record.

There are several cognitive theories supporting the idea that summative notes should be more effective for learning — that is, result in better recall — than verbatim notes. For example, the generation effect (Slamecka & Graf, 1978) refers to the finding that information is better recalled when it is self generated as opposed to being read or heard. Similarly, the levels of processing framework (Craik & Lockhart, 1972) also predicts that summative notes will be more effective for learning. This framework refers to the idea that recall improves as a function of the depth of cognitive processing: the more deeply information is processed, the better it is encoded and, the more likely it is to be successfully recalled. There are many ways of conceptualizing what depth of processing consists of (Dinsmore & Alexander, 2012), however, one way of thinking about these constructs comes from Jay et al. (2008), who argue that shallow processing involves encoding only superficial aspects of the stimulus, whereas deeper processing involves activating the semantic meaning of the stimulus (p.85). Verbatim note-takers are probably engaging in mostly surface level processing because they are attending more to transcribing the spoken words as accurately as possible, than to the meaning of those words. Summative notes require attention to the meaning of the spoken words and require more generative, deep processing. These considerations suggest that improved recall should be observed from summative notes compared to verbatim notes (especially when note-takers do not have the opportunity to review their notes).

Because summative notes require deeper cognitive processing, and because taking handwritten notes pushes many students toward summative notes, handwritten notes may indirectly be associated with better recall. Importantly, many formal note-taking strategies (e.g., clustering, Schultz & Di Vesta, 1972; concept mapping, Trochim, 1989;

and the Cornell system, Broe, 2013) focus on teaching students how to organize ideas through summative notes. Summative note-taking seems to be effective at improving recall relative to the typical note-taking strategies that students use (Makany, Kemp, & Dror, 2008; Slavina, 2018). In addition, handwriting also affords more opportunities for note takers to elaborate on information, by drawing, diagramming, or connecting disparate thoughts— all of which can be described as annotating. When note takers annotate (notes about their notes), they are engaging in a relatively deep-level of processing of the material and this should translate into better encoding and recall. While the deeper processing associated with handwriting gives greater opportunity to encode information for later retrieval, it also often comes at the expense of some information loss (sometimes referred to in the literature as "note quality") compared with verbatim notes. This information loss is most likely to be relevant when students review their notes for additional information.

Opportunity to Review. An important consideration in determining the effects of modality and note-taking style on recall is whether notes are reviewed. Note review is an additional opportunity for encoding information that was missed during initial note-taking that may result in improved recall. A large body of studies have investigated the role of note reviewing in recall (e.g. Fisher & Harris, 1973, 1974). As one might expect, these studies generally suggest that reviewing notes leads to better recall. For example, Kiewra et al. (1991) compared taking notes with review to only taking notes. They observed the highest recall among students who took and later reviewed their own notes. Similarly, Henk and Stahl (1985) performed a meta-analysis assessing the effects of reviewing notes independent of note-taking itself and found that reviewing one's own notes resulted in better recall than reviewing an instructor's notes. Kobayashi's (2006) subsequent meta-analysis compared note-taking and review with not taking notes or reviewing and found note-taking with review was significantly better.

When assessing recall, the available quantity of notes-taken are clearly related to what is able to be reviewed. For example, Kodaira (2017) randomly assigned 80 undergraduate students to take notes either by handwriting or by typewriting during a short (11 min.) lecture. After viewing the lecture and taking notes, students were asked to create a set of one-page summary notes for use by another student. This activity served as a review session. Finally, participants took a short multiple-choice exam on the lecture's content. Unsurprisingly, Kodaira found that typewriters were faster note-takers than handwriters ($d = 1.70$). Controlling for the quantity of the original notes, Kodaira also found a positive (but not statistically significant) benefit for typewriters on the quality of the original notes and a positive and statistically significant benefit for typewriters on the quality of notes produced for others. Controlling for the quality of the original notes and the quality of notes produced for others, Kodaira found a benefit for handwriters on the recall test. These contradictory findings are possibly explained by Kodaira's use of endogenous covariates in the last two models. That is, the model assessing the relationship between condition and the quality of summarized notes controls for a variable (the quality of the original notes) that may have been affected by condition. The same is true for the model assessing the relationship between condition and recall, which controlled for two potentially endogenous variables (the quality of the original notes and the quality of the summarized notes).

Note-takers' Intentions. Early theorists (DiVesta & Gray, 1972) proposed that note-taking serves two functions: (a) an encoding function and (b) an external storage function. The encoding function refers to the objective of gaining immediate improvement in attention, encoding, and recall from note-taking. The external storage function serves as a tool for later review, and, as discussed earlier, review provides additional opportunities to encode information and hence should lead to better recall. Kiewra (1989) noted at the time that this paradigm was the focus of nearly 100 studies (p. 148) when it was published and as of July 12, 2021 has over 650 citations in Google Scholar, suggesting that DiVesta and Gray's paper has been quite influential.

DiVesta and Gray's conception of the functional role of note-taking may be an oversimplification in the digital age. As computers have become more integrated into daily life, it is not hard to imagine that the typical typist's transcription speed is much faster than it was a generation ago (Karat et al., 1999; Dhakal et al., 2018). This increased transcription speed means that it is more feasible now than it was a generation ago to adopt the goal of creating a nearly verbatim record of the information — in other words, typical note takers now have a note-taking style option (verbatim notes) that was less realistic a generation ago. Of interest is that, when attempting to create a nearly verbatim record, note-takers likely engage in relatively little actual encoding. This could be an unintentional side-effect of cognitive overload — that is, the act of creating nearly verbatim notes might tax the cognitive system so much that there is little opportunity for encoding to take place. Or, this could be the result of a meta-cognitive decision on the part of the note-taker — that is, it could be the result of an intentional choice to focus on transcription now with the idea that encoding will occur later, during review or alternatively, that the information can be accessed later should the need arise. This latter point closely aligns with what Sparrow, Liu, and Wegner (2011) refer to as transactive memory (i.e., instead of a memory of needed information, a memory of where needed information can be found). Either way, we would expect less encoding among note-takers who take verbatim notes. Slavina, (2018, p.40) demonstrated this directly with a note-taking experiment in which participants saved their notes into folders and were able to recall "easy" to remember facts better than where they saved the information but for "hard" facts they better remembered where they saved them (which folder) this was consistent across note-taking modalities.

Recall Measurement. The most common dependent variable in studies assessing the relative advantages of handwriting vs. typewriting notes is recall, typically measured as performance on a short quiz or test on the content learned. The implication of the focus on recall is that learning is operationalized as information stored in and—more importantly—accessed, or recalled, from long term memory. The assessment of recall varies in important ways across studies. One dimension along which studies vary is the availability of an opportunity to review the notes, as discussed in the previous section. In this section, additional ways that recall measurement varies across studies and why these measures might affect outcomes are discussed: timing of assessment, presence of a distractor task, length of assessment delay, and the knowledge dimensions tested (e.g., factual vs. conceptual).

Timing of Assessment. Some modality studies assess learning right after the presentation of the learned material while others assess learning either after a delay (usually one day to one week) or immediately after presentation of the learned material but with a distractor task. The timing of assessment matters from a cognitive standpoint because information is initially stored in temporary systems (sensory memory, working memory, and short-term memory), all of which have significant constraints with respect to the amount of information that can be held and the length of time information can be held without rehearsal.

Distractor Task. Because rehearsal can be used to keep information active in the temporary memory systems, researchers often build in distractor tasks in an attempt to assess immediate long-term memory recall shortly after initial information acquisition. The idea is that participants perform a task that taxes their temporary memory systems. One example is the complex span task, which requires participants to simultaneously perform a memory task (e.g., remembering a series of letters) and a processing task (e.g., determining whether a simple equation is correct). Taxing working memory in this way should reduce the likelihood that the information relevant to the note-taking activity remains active in working memory. This should more accurately assess recall of information encoded in long-term memory.

Length of Assessment Delay. The length of time between when information was recorded in notes and when it is recalled also varies between studies from minutes to weeks. Length of assessment delay is important because assessment immediately after note-taking might provide an instructor with immediate feedback about the extent to which students were able to encode the information that was presented while long term delays might inform how well it was retained. Stated differently, immediate recall without a distractor task and delayed recall are assessing somewhat different things (encoding and short-term memory recall vs. long term memory recall).

Knowledge Dimensions Tested. The questions typically used to assess recall can roughly be grouped into two categories: factual and conceptual. Factual questions measure the presence of specific applicable information, whereas conceptual questions measure the understanding of the interrelationships between constructs and other information. It is worth noting that the knowledge dimensions tested will often be confounded with the nature of the test. That is, studies assessing recall of facts will often do so using multiple choice questions, whereas studies assessing recall of concepts will typically use short or long answer questions. This observation is important because it implies that if, for example, one note-taking modality appears to be superior to another note-taking modality for a particular knowledge dimension, it will be hard to interpret that interaction (because knowledge dimension tested is likely to be confounded with the nature of the test).

The Fundamental Problem of Note-taking Modality Research

In the sections above I provide an overview of a few theoretically important moderators of the note-taking modality effect. Many of these characteristics are not discussed in most studies. Furthermore, the fundamental problem of this line of research is that when one manipulates modality, other things covary with that manipulation even if this confound is not explicitly acknowledged. For example, note takers using typing input will often attempt to take verbatim notes (recall that Mueller and Oppenheimer [2014] concluded that instructing typewriting note-takers not to take verbatim notes was "completely ineffective", p. 5). If verbatim note takers are expected to encode less, then they can be expected to have worse recall in situations in which they do not have access to their notes, for example, on a quiz immediately after lecture. Gong and Rodd (2020) also provided some evidence for this idea. These authors manipulated two factors: modality (handwriting vs. typewriting) and expectations for later note access (notes will be available later vs. not), which is one way of manipulating note-taker intentions (i.e., an expectation that notes will not be available for later review before a test, would probably lead note takers to engage in more encoding while taking notes). Recall was assessed after a one week delay. The authors found a significant modality by access expectation interaction, such that the recall advantage for handwriters entirely disappeared among participants who were told that notes would not be available for review, suggesting "that the extent to which participants engage in cognitive offloading depends on the modality of their note taking" (p. 17). Stated differently, when researchers manipulate whether or not participants expect to have access to their notes for review, they might also — intentionally or not — manipulate the extent to which participants engage in other behaviors such as encoding and offloading.

Overview

Most people take notes in some form, often to improve recall but sometimes for other reasons (e.g., to create a record of how to find information, or to help maintain attentional focus). Many studies have examined the recall benefits of handwriting versus typing notes. From a theoretical perspective, there are a number of potentially important moderators of any modality effect, including the note-taking style, speaker's pace, the note-taker's transcription speed, and the measurement of recall. This study uses systematic reviewing and meta-analysis of studies examining the effect of note-taking modality on recall among secondary and postsecondary students. This review is important because the research findings — including previous reviews — on note-taking modality are understandably equivocal, as previously mentioned, because there are several factors, like note-taking style, that can potentially confound the relationship of interest. Meta-analysis is especially useful in this environment, because synthesizing effects observed in many different studies will increase statistical power, provide a more generalizable result, and, the examination of potential study-level moderating variables can create a platform upon which further research can build. Furthermore, I will collect information on theoretically important moderators (see Table 1) to document how often these are discussed in note-taking modality studies (see Table 2), and will attempt to identify potential confounds such as the extent to which note-taking modality assignment leads the study participants to engage in other behaviors (such as a particular note-taking style).

CHAPTER III

METHODS

Inclusion and Exclusion Criteria

To be included in the meta-analysis, studies had to meet several criteria. Specifically, studies had to manipulate note-taking modality (i.e., assign participants to take handwritten or typed notes), include an external comparison group (formed using randomization or similar process), and assess information recall. Recall accuracy could be assessed using a variety of means including multiple choice, true/false, and short answer tests administered post transcription. Distractor tasks were not required, but their presence or absence was coded (see Table 2.3). Writing must have been done using pen and paper. Typing must have been done using a hardware keyboard.

Furthermore, studies had to be conducted in a secondary or higher education classroom or laboratory setting involving participants who served as learners for content delivered verbally (perhaps in conjunction with written material, such as presentation slides). Therefore, studies of note-taking in criminal or civil justice settings were excluded, as were studies that involved taking notes from text. Studies involving logographic based languages were excluded (e.g., Chinese, Japanese, and Korean) because of the varied graphomotor movement and transcription effort compared with letter based languages. There were no publication status, age, time period, geographical, or cultural restrictions. I assessed reports in languages other than English using Google Translate.

Literature Search Strategies

Electronic Database Search

I developed the electronic database search strategy in consultation with a research librarian. The ProQuest and EBSCO search platforms were used to search the databases likely to contain relevant research (see Table 3). I developed and ran the search queries on each platform to find articles that discussed the primary constructs of interest for this systematic review (viz., handwriting, typing, and recall). See Table 4 for the specific queries that I used, including different variants on the key constructs[3]. Because the recall concept block has the potential to limit the retrieved studies that use terms relevant to recall, I tested the search against ten studies that I knew to be eligible for this review (see table Table 5). All ten studies were identified by the search. Also, I used the stringdist package (van der Loo, 2014) with the statistical software R (R Core Team, 2021) to compute a similarity score between abstracts and identify duplicates of the articles already found from a previous pilot literature search.

Ancillary Search Strategies

On 5 June 2021, I searched the Registry of Efficacy and Effectiveness Studies (https://sreereg.icpsr.umich.edu/sreereg/) and the Open Science Registry (https://opensciencereistry.org) using the following terms: "note-taking", "notetaking", "note taking". I used Twitter to solicit studies that might not be publicly available (e.g., in press articles and articles that were not submitted for publication), mentioning Division

[3] I found that “note-taking” was equivalent to "note taking" across both of the search platforms used and I elected to use the Oxford English Dictionary suggested form of "note-taking".

15 of the American Psychological Association (educational psychology; @APADiv15), the American Educational Research Association's Writing and Literacy Special Interest Group (@writinglit), and using a generic hashtag (#edpsych). I also solicited studies that might not be publicly available from the Educational Psychology interest group on Reddit. In addition, for all articles that were eligible for inclusion in this systematic review, I searched through the references of the papers, reached out to authors with multiple included studies, and used Google Scholar and Web of Science to perform forward citation searches of two seminal papers (Di Vesta & Gray, 1972; Mueller & Oppenheimer, 2014).

Screening Process

9,063 abstracts were downloaded from electronic database searches. Similarity scores were computed and scores > 2 mad from the median were removed as duplicates. The final number of abstracts was 1235.

Title and Abstract Screening

Potentially eligible documents were reviewed independently by two trained screeners using the screening questions in Table 5; a pilot round of screening was then used to test the screening questions and process. The goal of this process was to eliminate from consideration all documents that were clearly not eligible for review. After the pilot round the two reviewers agreed on 99% (8 conflicts in abstract review) of abstracts (Cohen's Kappa = .99). Disagreements were resolved in consultation. A total of 1,129 of the 1,235 (91%) documents were determined not to be eligible for review.

Full Text Screening

I attempted to obtain all documents that were not eliminated during the title and abstract screening phase. Of the 106 documents that were retained, I was able to locate 103 of these, the full text of the documents were screened and 33 (32%) were determined to be eligible for review.

Coding Procedures

All studies were coded by two trained researchers working independently. Disagreements were resolved by discussion. Coding involved extracting the study context, sample, nature of the intervention, outcome characteristics, and effect size information from the full text documents that were eligible for review. Where there was missing effect size information for studies less than 15 years old, an attempt was made to contact the study authors. Table 2 is a meta-table describing the content of several tables used to describe the studies included in the review.

Statistical Methods

I used meta-regression with inverse variance weights to conduct meta-analysis of effect sizes examining the effect of modality on recall. Meta-regression is a statistical technique that treats individual effect sizes as observations and regresses them, along with relevant study covariates, on the outcome of interest. Meta-analysts have to choose a statistical model for both the overall analysis and for any moderator analyses. For the overall analysis, I chose a random effects meta-regression model because the differences in populations, methods, and study design countered the theoretical assumption of the fixed effect model that each effect size estimates the same population parameter and therefore that their differences are the sole product of sampling variability (Hedges & Vevea, 1998). Rather, it seemed likely that each study had additional sources of variability in the study characteristics associated with effect sizes. For the moderator analysis, I chose a mixed effects model which involves using random effects meta-analysis to synthesize effects within levels of a moderator and then fixed effects meta-analysis to test the difference between the levels of the moderator.

Because there was little consistency in scaling of the outcome variable across studies, Hedges' *g*, the standardized mean difference (*d*) corrected for small-sample bias, served as the effect size of interest for this meta-analysis. Effect sizes and their standard errors were computed using standard formulas (Borenstein & Hedges, 2019) from functions built into the metafor (Viechtbauer, 2010) package in R (R Core Team, 2022). When needed, test statistics and other information (such as an independent groups *t*-test and degrees of freedom) were converted to Hedges' *g*. Specifically, when study results were presented in dichotomized fashion, e.g., percentage correct on a quiz, Hedges' *g* was computed using custom R functions derived from formulas presented in the What Works Clearinghouse's *Standards and Procedures Handbook version 5.0*, (2022; see formulas starting on p.177). Effect sizes were transformed from test statistics using the compute.es (Re, 2013) package in the R software. Heterogeneity in effect sizes were assessed using Cochran's *Q*.

The data structure of this study complex, with effect sizes (88) nested in samples (42) which are nested in reports (33), and the proportion of shared variance at the report level (i.e., intra-class correlation) is .55. This means that analyzing the data as if the effect sizes were independent would result in a violation of the assumption of statistical independence, if I were to use traditional meta-analytic techniques. To address this concern I used a multilevel random-effects meta-regression model. I also adopted analytic strategies that were designed to minimize type I error, these included using restricted maximum likelihood (Harville, 1977) estimation, cluster robust tests and confidence intervals (Viechtbauer, 2015), bias-reduced linearization adjustment (Bell & McCaffrey, 2002), Hotelling's T-squared reference distribution (Hotelling, 1931), and Satterthwaite (1946) adjusted degrees of freedom.

All systematic reviews should involve an assessment of study quality (Applebaum, et al., 2015). Study quality indicators are factors influencing the believability of statements arising from the study regarding the causal relationship between modality and recall. An indicator approach, or drawing consensus from attempting multiple methods, to study quality was used instead of a recognized study quality scale because study quality is (a) not captured well with scale summary scores because it is a multidimensional construct, and (b) a contextually dependent construct, and no validity scales have been proposed that address note-taking modality research (Valentine, 2019). The individual study quality indicators chosen were (a) the use of random assignment to place participants into different modalities, (b) the overall attrition rate, and (c) the absolute value of the differential attrition rate (Valentine, 2019).

For the final statistical model, I planned to predict recall from modality, controlling for the following study-level characteristics: the speaker's pace in WPM, whether the note-taker's transcription speed was used as a control variable, whether note-takers had the opportunity to review their notes, the note-taker's intentions (i.e., whether note-takers expected to have access to their notes before recall was evaluated), how recall was measured (e.g., conceptual vs. factual questions), whether there was a delay in recall assessment[4], and the three indicators of study quality. However, as elaborated in the results section, this full model was not viable due to missing data and a lack of variation across studies for some of these dimensions (see Table 2.3).

Possible data censoring through possible publication bias or selective reporting were assessed with the rank correlation (Begg and Mazumdar, 1994), Egger's test (Sterne & Egger, 2005), and excess significance tests (Ioannidis & Trikalinos, 2007), along with the Henmi and Copas approach (Henmi & Copas, 2010), and a funnel plot with trim and fill (Duvall & Tweedie, 2000). Multiple publication bias methods, measures for assessing the potential of observing only the results of statistically significant findings in publications and the censoring of non-significant findings, were used due to the different assumptions each method uses so that I could triangulate across methods (Banks, Kepes, & McDaniel, 2012).

Data wrangling was performed using the dplyr (Wickham et al., 2022) and stringr (Wickham, 2019) packages in the R software (R Core Team, 2022). All statistical analyses were done using the metafor package (Viechtbauer, 2010) and R. Outlier analysis was performed using studentized residuals (Viechtbauer & Cheung, 2010) and flagged effect sizes were marked (see Table 6). Influence analysis was conducted using the DFBETAS, Q test statistics, and studentized residuals by leaving each effect size out of the analysis and observing the change in overall effect size estimates, heterogeneity, and model fit along with evaluating the cause of their influence and their potential for highlighting additional moderators. I used the default settings for influence diagnostics in the metafor package to identify potentially influential studies (Viechtbauer, 2022, see p.

[4] I initially neglected to mention recall measurement in the methods section of the dissertation proposal, but listed it in the theoretically important potential moderators section of the introduction.

104) and one study (Cubilo, 2017) was flagged as potentially influential. I chose to leave the study in the analysis. See Appendix A for a link to all of the code contributing to the analyses reported in this manuscript.

CHAPTER IV

RESULTS

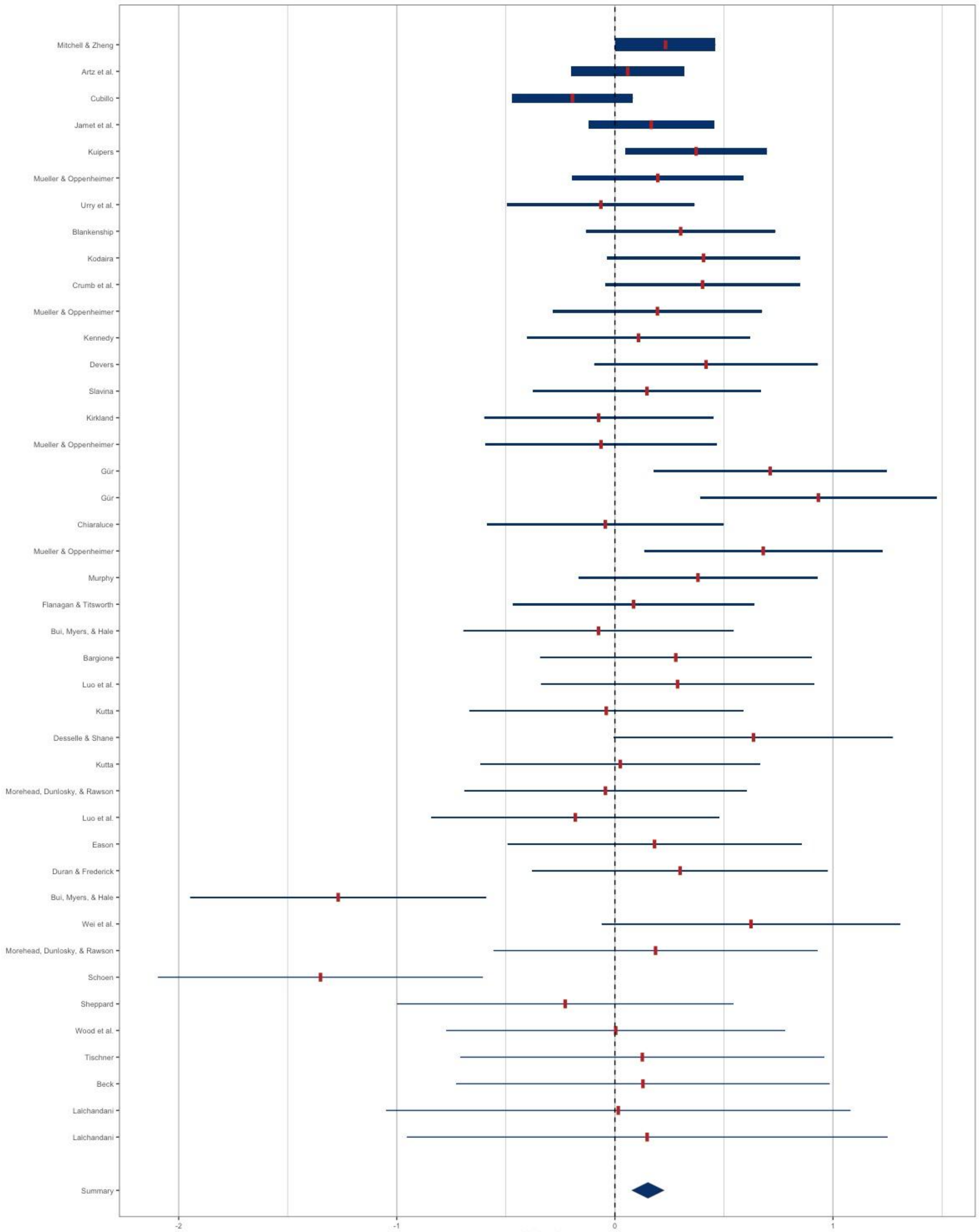
Description of the Included Studies

As can be seen in Table 2.2, the typical study in this review was conducted on a convenience sample of American undergraduate students who were randomly assigned to a note-taking modality. In general, students were exposed to a short (10-15 minute) lecture on a topic with which they had little familiarity. Subsequently, they were administered a short quiz developed by the study authors — the psychometric properties of which we know very little — immediately after the lecture and/or after a delay of 2-7 days. Most immediate post-tests were preceded by a distractor task (see Table 2.3).

As an overview of the data contributing to this meta-analysis, I developed a forest plot. Forest plots display study results (effect sizes and confidence intervals) in a compact way. I used the aggregate function in R's metafor package to create a single, overall average effect size for each independent sample in order to produce this plot (see Figure 1). I used confidence interval line thickness to indicate each independent sample's weight (thicker lines indicate greater weight). I also sorted the plot by the standard error, so each independent sample is ordered from the most precise to the least precise estimates. As can be seen, two independent samples have larger negative effect sizes than is typical but their standard errors overlap with the standard errors of many of the smaller sample studies, and the outlier and influence analyses did not suggest that these were overly influential observations. At the effect size level (data not shown), there were there were 76 effect sizes (13 of which were statistically significant) indicating a benefit for handwriters, and 20 (three of which were statistically significant) effect sizes indicating a benefit for typewriters. One effect size was exactly zero.

Figure 2

Forest plot of independent sample effect sizes ordered by standard error



Note. Each independent sample used in this review is represented in the forest plot. Circles indicate the average effect size within each independent sample, and the lines extending out from the circles indicate the 95% confidence interval for each effect size.

Initial Data Analysis and Results

Below I present three different approaches to data analysis. First, I present the results for an intercept-only model, which is equivalent to the random effects weighted overall average effect size (Table 7). Next, I present meta-regression results for the three univariate moderator tests that I committed to *a priori* and that had sufficient data to support analysis (Table 8). Finally, I present the results of the combined meta-regression model that tests all three moderators simultaneously (Table 9).

The intercept model estimates a statistically significant and positive overall effect for handwriting compared to typewriting across all studies. Specifically, the overall meta-analytic effect size is g = +0.144 [0.023, 0.265] indicating a small, statistically significant, positive benefit in recall associated with handwriting over typewriting. To interpret this effect, consider a study that administered participants a quiz and reported results in terms of the percentage of items that were answered correctly. The median percentage correct from my dataset is approximately 50% in the typewriting condition. If the participants who took notes by typewriting scored 50%, the effect size of g = +0.144 implies that students taking notes by hand would have scored 56.5% on average (see Valentine, Aloe, & Lau, 2016, for a description of this and other methods for translating effect sizes into more interpretable metrics).

The homogeneity test was statistically significant, $Q(87) = 139.6$, $p < .001$, $\tau^2 = .080$, $I^2 = 55.9\%$, suggesting more variability in effect sizes than would be predicted by sampling error alone. Unlike confidence intervals, which describe the likely distribution of the population effect, predictions interval describe the likely range within which some future observation might fall (e.g., the next study) and therefore represent a good way of thinking about the magnitude of true heterogeneity (Borenstein et al., 2016). For this overall effect size, the prediction interval ranges from a low of g = -0.013 to a high of g = +0.301. Using the method described above articulated by Valentine et al. (2016) to contextualize this overall effect size, the prediction interval suggests that the likely range in which the next study conducted on the effects of note-taking modality might find anything from a tiny benefit favoring typewriting (if the typewriting group scored exactly 50% the lower limit of the prediction interval suggests that the handwriting group would score 49.4%) to a rather large benefit for handwriting (if the typewriting group scored exactly 50% the upper limit of the prediction interval suggests that the handwriting group would score 63.3%), suggesting a non-trivial degree of heterogeneity in the true effect sizes.

Outlier and Influence Analysis

The outlier and influence analysis found three potential outliers in the sample of 88 effect sizes. No effects were removed from the analysis because the magnitude of influence was insufficient to remove them; see Figure 2 for a visualization of this at the study level and Table 6 for outlier and influence diagnostics.

Moderator Analyses

In an attempt to explain why true effect sizes might differ from study-to-study, I planned on conducting several univariate and one multivariate analysis. Due to the way the studies were conducted (i.e., data were unreported and therefore missing, or because there was little variation on the factor across studies) only three of the originally proposed factors were feasible for univariate modeling (viz., delay, review, and measure type) and including all three of these factors in the same model reduces the number of studies from 33 to 14 and samples from 42 to 19, introducing potential for finding statistical significance as a byproduct of the sub-sampling through missingness (see Table 9). The variables that could not be included in the planned univariate tests are reported and described in Table 6. For completeness I conducted additional exploratory univariate analyses based on factors that I did not commit to examining in advance of data collection. These analyses are reported in Table 11.

Each of the planned univariate models that I could run used a different subset of studies due to the varying missingness of each variable in the data. None of these planned univariate moderator effects were statistically significant (see Table 9). In the planned multivariate model, the only statistically significant moderator was the opportunity to review. The standardized mean difference effect size of g = -0.414 for this variable suggests that the advantage for handwriters is nearly canceled out when there is an opportunity to review. That is, among studies that allowed participants to review their notes before taking the post lecture quiz, if typewriters scored 50% then handwriters scored 52% on average, controlling for the other variables in the model. However, it should be emphasized that compared to the intercept-only model, the sample of studies is significantly reduced, and in addition this moderator was not statistically significant in the univariate model (though the substantive interpretation of the results is similar across models). These results should be interpreted with caution.

Publication Bias Tests

Almost half (48%; 6 dissertations, 8 theses, 2 conference presentations) of the 33 studies included in this review were unpublished, suggesting that publication bias is unlikely to be a strong threat to the validity of my conclusions. In fact, four of the five publication tests that I committed to examining suggested that there is little evidence of publication bias[5]. The rank correlation test (p = .43) and Egger's regression test (p = .39) were both non-significant. The trim and fill procedure did not identify any possibly censored effect sizes (see Figure 2; made with the metaviz R package, Kossmeier et al., 2020), and the Hemni and Copas meta-analytic estimate changed very little (+0.002 standard deviations). Only the test of excess significance suggested some possibility of publication bias (p = .03). Given the relatively large percentage of unpublished studies and the near-agreement among the publication bias tests, it seems unlikely that publication bias is operating in an important way in this meta-analytic dataset.

Figure 3

Funnel plot with trim and fill

[5] These tests were run using the aggregate function in the metafor package to create a single effect size per study because they were not designed to handle the assumption of independence being violated.

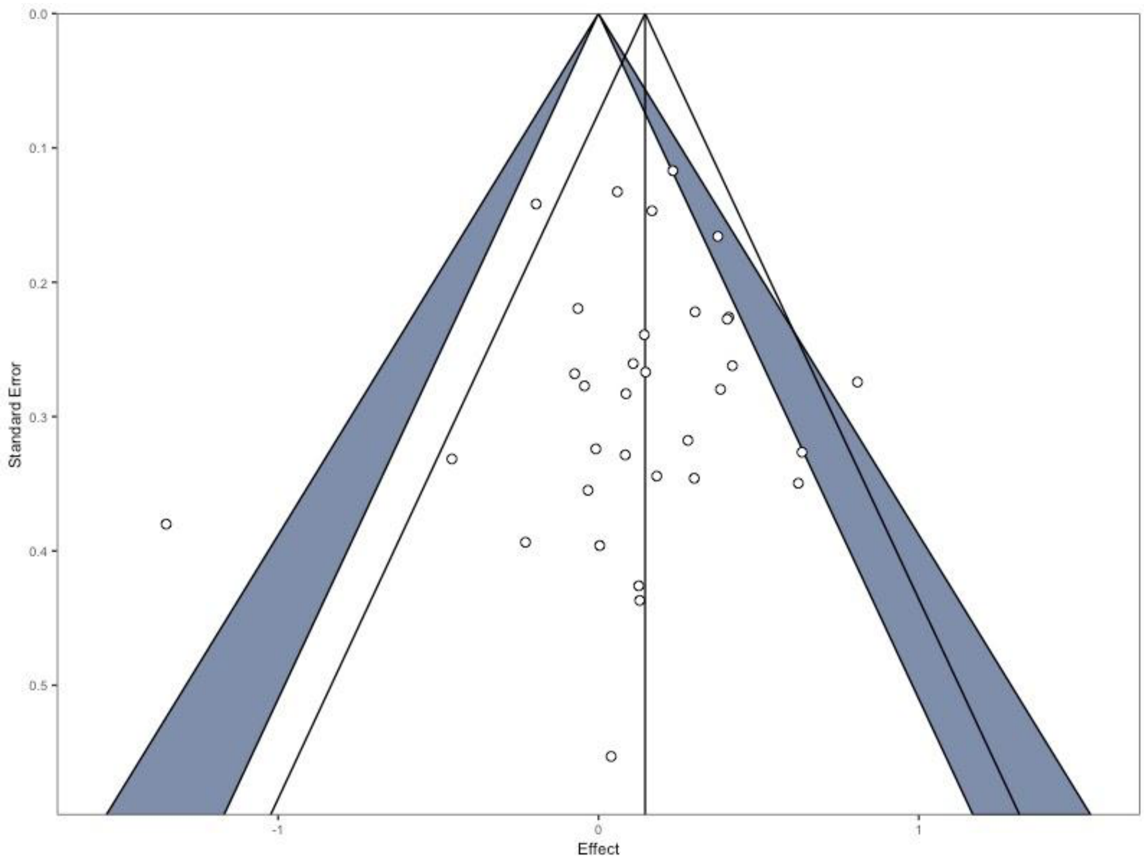


Figure note. Funnel plot showing each study’s aggregated effect size.

CHAPTER V

DISCUSSION

This systematic review and meta-analysis involved 33 reports, 42 independent samples, and 88 effect sizes, all evaluating whether there are recall differences — usually operationalized as scores on a quiz given after exposure to a short lecture — between participants taking notes by handwriting vs. typewriting, that is, the modality effect. Effect sizes were nested in participants and participants were nested in studies, and treating these effects as independent would violate the statistical assumption of independence, leading to standard errors that are too small. Therefore, a robust multilevel random effects meta-regression model was applied to address the dependence in the observed effect sizes. A statistically significant overall meta-analytic average was found $g = +0.144$ [0.023, 0.265], $p = .021$, benefiting handwriters over typewriters. Given that the typical quiz average was about 50% percent correct in the typewriting condition, this effect translates to an average percent correct of about 55.9% in the handwriting group.

The test of homogeneity was statistically significant, suggesting the studies vary more than sampling error alone would predict. Therefore, in an attempt to explain this heterogeneity, I ran both univariate and multivariate models to assess the potential impact of moderators both individually and simultaneously. It should be noted that most moderator analyses I planned on conducting could not be run due to either little variation on the dimension or to pervasive missing data. However, none of the viable moderator analyses explained the observed heterogeneity, that is, none of the three moderators were statistically significant when tested using univariate models. When tested simultaneously (i.e., the multivariate model), opportunity to review was statistically significant and, in fact, almost eliminated the advantage for handwriters. However, this analysis is based on a much-reduced and potentially unique subset of studies, and it is possible that it is simply an artifact.

This systematic review and meta-analysis provides some evidence that the opportunity to review one's notes moderates the effect of note-taking modality on recall. Both note-taking modalities benefit from note-taking review because it primes recall of previously stored information. In addition, note-review is another opportunity to encode missed information. This theoretically puts verbatim note-takers at an advantage because of additional detailed information that would otherwise be summarized and because verbatim note-takers engage in less encoding when taking notes than summative note-takers. So as discussed in detail in the Introduction, it was expected that the opportunity to review notes would moderate the modality effect on recall.

It is worth noting that some studies used notes-review as a factor, allowing me to examine the effects of asynchronous review. Of note is that no studies have attempted to control synchronous note-review, that is reviewing notes while taking them. This is possible to do with handwriting for example by using digital e-writers, invisible ink, or ink that matches the color of the writing paper. Similarly, synchronous note-review can be tested among typewriters by not allowing users to see what they are typing. However, visual feedback is fundamental to the note-taking process and eliminating the opportunity for synchronous review would likely affect note legibility and interpretability. Critically eliminating the opportunity for synchronous review would likely increase the cognitive burden on note-takers. These problems would likely make it more difficult to interpret such studies.

I also examined the effect of using relatively immediate vs. a short (2-7) days delay on the modality effect. Because memories decay over time, if delay in recall measurement is a factor in a primary study, we would expect that there would be a main effect for delay, such that participants assigned to take a test immediately after exposure would score higher than participants assigned to take a test after some delay. The main effect for delay in recall theoretically interacts with note-review, because note-review can replenish lost information. So, for example, if both note-review and delay were factors in an experiment, we would expect to observe an interaction between these factors, such that the effect of delay would be smaller in the presence of note-review. However, I did not observe a significant effect for delay in these studies. It is possible that the delays observed in these studies were too short, or that the amount of information to be learned was too insubstantial (typically short 10-15 minute lecture videos), for the delay effect to be observed.

Finally, I examined the nature of the questions asked on the recall quiz as a potential moderator. The factual vs. conceptual question type distinction is a factor often called out in modality studies as a plausible moderator. But there is inconsistency with how the distinction is applied, both within studies and at the synthesis level. Within studies, most quizzes administered in studies in this review were developed by the study authors, and we know very little about their validity. We also know relatively little about the consistency with which authors across studies applied the factual and conceptual labels to the questions on their quizzes. At the synthesis level, I relied on the authors' labels when they were provided, but I also attempted to categorize tests based on information provided in the studies and this process might have introduced error as well. In addition, I could only place 72% of the effect sizes into factual or conceptual question categories. It is possible that these considerations contributed to the very small, non significant effect sizes that I observed for this potential moderator.

Limitations and Directions for Future Research

In the following sections, I discuss three limitations and consequent directions for future research arising from my systematic review. These are: the extent to which the studies included in this review are practically relevant, the extent to which the studies included in this review are theoretically relevant, and the fact that there was a non-trivial degree of effect size heterogeneity that was largely unexplained. I conclude by offering recommendations to note-takers based on the results of this systematic review.

Practical Applicability of the Evidence

Scholars sometimes use the term "applicability" (a type of external validity) to describe the extent to which the conditions in experiments apply to real world conditions. The studies in this systematic review are not overly applicable. In most studies, students experienced a single 10-15 minute video lecture on unfamiliar content, took a short quiz developed by the study authors quickly after the lecture or, at most, a week later, and may or may not have had an opportunity to review their notes. In contrast, of course, many real-world applications of note-taking involve much more content, at least some familiarity/prior-knowledge of that content, knowledge-related applications that extend beyond recall (for example, synthesis), and over a longer time frame. All of this suggests that while the effect size observed in this systematic review and meta-analysis is probably true, we do not know much about how these findings, based on a set of narrowly designed studies, will translate into a more practical context.

Theoretical Applicability of the Evidence

Another issue relates more to the theoretical meaningfulness of these results. Specifically, it is unclear if the statistically significant overall meta-analytic average observed should be taken at face value and attributed to modality. From a theoretical perspective, these studies are largely beside the point, because few studies attempted to break the connection between modality and note-taking style and this design feature is needed to approach answering the theoretical question about which note-taking modality is superior for learning. This is because when researchers assign participants to a note-taking modality they are also indirectly assigning them to different note-taking styles (i.e., there is a strong tendency for typewriters to take verbatim notes whereas handwriters are more likely to take summative notes), and these different styles have inherent differences in their effects on cognition and therefore recall. Note-taking modality pairs and complements style through the interplay of transcription capacity and speaker pace. When speaker pace is fast (such as is typical in a university lecture) and transcription capacity is slow (handwriting; unimanual transcription) students are less likely to attempt to take verbatim notes, but when transcription capacity can approach speaker pace (typewriting; bimanual transcription) there is incentive to transcribe as much as possible (verbatim) for future storage without thought for the potential short term cost in encoding and comprehension that accompany this note-taking style.

In this regard, only two studies (out of 33) examined note-taking style as an experimental factor (Bui et al., 2013; Mueller & Oppenheimer 2014), so I was unable to test this as a potential moderator. In Bui et al. (2013), there was a small but not statistically significant advantage for *typewriters* (g = +0.015) when participants were constrained to take summative notes. But, in Mueller and Oppenheimer (2014; it should be noted that the authors identified fidelity of implementation issues with the verbatim condition, p. 5), the advantage for handwriters over typewriters persisted even when participants were assigned to take summative notes. However, this advantage was somewhat smaller (g = -0.04) relative to the handwriting advantage observed when note-takers were not assigned to a particular note-taking style. These results suggest that one potentially illuminating direction for future research will be to examine the note-taking style as part of the dynamic system and environment that produces the kinds of notes taken with different modalities. That is, future studies could contribute to the theoretical discussion regarding modality superiority by including note-taking style in the study design, for example by training note-takers on strategies for taking verbatim notes and summative notes using a computer and keyboard, and then randomly assigning trainees to a note-taking style. Such a study should provide a fair test of the theoretical considerations suggesting that summative note-taking is superior to verbatim note-taking because it would hold modality constant. In addition, there is a strong need for more studies in more naturalistic settings/conditions, using materials that students are actually responsible for learning, multiple exposures to different parts of the material, and tests that are relatively far removed from content exposure. Finally, to help establish a solid basis for making practical recommendations, future studies should also be designed to address the extent to which speaker pace, student transcription capacity, working memory, and student prior topic knowledge factor into the impact of modality.

Unexplained Heterogeneity

An issue that limits the ability of this evidence base to contribute to both theory and practice relates to the unexplained heterogeneity observed in these results. There is good reason to suspect that the modality effect should be moderated based on theory — that is, that the main effect of modality should interact with other factors, such as speaker pace. There is potential evidence from this study that the opportunity to review does in fact moderate the modality effect: When the factor for whether or not participants had an opportunity to review their notes was included as a study-level moderator along with two other potential moderators (whether the quiz was given on the same day as exposure to the lecture and whether the quiz was based on factual or conceptual questions), the overall average dropped from about $g = +0.47$ when there was no opportunity to review to about $g = +0.05$ when there was an opportunity to review, controlling for the other two variables in the model. This finding does make some conceptual sense — if typewriters have notes approaching a verbatim transcription of the lecture material and then are allowed to study these prior to taking the quiz, this should help make up for the lack of deep processing while the notes were being created, and it is in line with the research on the effects of reviewing notes introduced earlier. There are two important caveats to this finding. One is that opportunity to review was not a statistically significant moderator when tested by itself in a univariate model ($p = .091$), though, to be fair, the story that emerges from the univariate analysis is the same as the one that emerges from the multivariate model, that is, that having an opportunity to review notes largely eliminates the recall advantage for handwriters over typewriters. Perhaps more important is that the multivariate result is based on a potentially unique subset of studies that had no missing information on all three potential moderators I was able to test, and therefore is based on less than half of the available evidence (i.e., 14 reports out of a possible 33). Clearly, one important direction for future research might be to treat the opportunity to review as an experimental factor that is manipulated along with note-taking, providing a stronger test of this potential effect.

Regardless of whether one believes that opportunity to review is operating as a moderator in these studies, accounting for it reduced but did not eliminate unexplained effect size heterogeneity. This unexplained heterogeneity adds uncertainty to the interpretation of the evidence base — essentially it implies that the effect sizes vary more than would be expected given sampling error, but given the current data I cannot fully describe why this happened.

Directions for Future Research

Looking forward, with the increasing utilization of on-line learning, it is unclear what lecture note-taking modality research should look like. Lecture note-taking research has previously operated under the paradigm that externally there is limited exposure to information, with note-taking operating as one solution to the fact that there is a limit to how much information can be encoded at once. That is, note-taking reduces the amount of information that would otherwise be irretrievably lost. It is relatively easy to produce transcriptions of recorded material, so the fact that more instruction is occurring asynchronously suggests that students should embrace note-taking strategies that promote qualitatively better information processing. This suggests that it may be productive for researchers to emphasize different strategies such as summarizing and annotating a transcript and note-taking from text.

Recommendation to Note-Takers

From a practical standpoint, the motivating question for this systematic review and meta-analysis can be thought of as this: If a student asks for advice on how to take notes, by using handwriting or typewriting, how should we respond? The answer, of course, is that "it depends". If the reason for taking notes is to pass a quiz or test shortly after lecture, then the results of this review suggest that handwriting is the better option that, on average, will result in a somewhat better score. This is probably because the student will likely take summative notes by hand, and even though I could not test this directly, there are strong theoretical reasons to believe that summative note-taking results in deeper processing and therefore better encoding of the information. However, this effect is small and performance will likely be improved by factors that the student is likely already aware of, such as limiting distractions while taking notes. In addition, if there is an opportunity to study for the quiz or test, then performance will likely be enhanced by following advice from the educational psychology literature (e.g., Dunlosky et al., 2013). Finally, if the student is taking notes to pass a test at some point beyond the near-term, for example a midterm exam, then the answer to this question will have to wait until a collection of studies designed to address this question are available.

REFERENCES

* studies included in the meta-analysis

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73*(1), 3-25. http://dx.doi.org/10.1037/amp0000191

Arif, A. S. & Stuerzlinger W. (2009). Analysis of text entry performance metrics. *2009 IEEE Toronto International Conference Science and Technology for Humanity*, pp. 100-105. https://doi.org/10.1109/TIC-STH.2009.5444533

Ash, P. & Carlton, B. J. (1953). The value of note-taking during film learning. *The British Journal of Educational Psychology 23*(2), 121-125. https://doi.org/10.1111/j.2044-8279.1953.tb02848.x

Arif, A. S., Stuerzlinger, W. (2009). Analysis of text entry performance metrics. *2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH)* (pp. 100-105). https://doi.org/10.1109/TIC-STH.2009.5444533

*Artz, B., Johnson, M., Robson, D., & Taengnoi, S. (2020). Taking notes in the digital age: Evidence from classroom random control trials. *The Journal of Economic Education, 51*(2), 103-115. https://doi.org/10.1080/00220485.2020.1731386

Bain, K., Basson, S. H., & Wald, M. (2002) Speech recognition in university classrooms. *Proceedings of the fifth international ACM conference on Assistive technologies*, 192-196. https://doi.org/10.1145/638249.638284

Banks, G. C., Kepes, S., & McDaniel, M. A. (2015). Publication bias: Understanding the myths concerning threats to the advancement of science. In C. E. Lance & R. J. Vandenberg (Eds.), *More Statistical and Methodological Myths and Urban Legends* (pp. 36–64). Routledge. https://doi.org/10.4324/9780203775851

*Bargione, J. L. (1987). *Microcomputers and recall: effects of typing ability and quality of notes* (Publication No. 303637077)[Doctoral Dissertation, University of Southern Mississippi]. ProQuest Dissertations & Theses Global.

Barbash, F. (2014, April 28). Why students using laptops learn less in class even when they really are taking notes. *The Washington Post.* https://www.washingtonpost.com/news/morning-mix/wp/2014/04/28/why-students-using-laptops-learn-less-in-class-even-when-they-really-are-taking-notes/

Bartolomeo, P., Bachoud-Lévi, A. C., Chokron, S., & Degos, J. D. (2002). Visually-and motor-based knowledge of letters: Evidence from a pure alexic patient. *Neuropsychologia, 40*(8), 1363-1371. https://doi.org/10.1016/S0028-3932(01)00209-3

*Beck, K. M., Hartley, J. S., Hustedde, S. L., & Felsberg, T. C. (2014). Note Taking Effectiveness in the Modern Classroom. *Educational Psychology, 77*(5), 522-532.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*, 1088-1101. https://doi.org/10.2307/2533446

*Blankenship, M. W. (2016). *Comparing note taking and test performance in methods and modes of note taking conditions* (Publication No. 1822245526) [Doctoral Dissertation, Ball State University]. ProQuest Dissertations & Theses Global.

Bledsoe, D. (2011). Handwriting speed in an adult population. *ADVANCE for Occupational Therapy Practitioners 27*(22), 10.

Borenstein, M., & Hedges, L. V., (2019). Effect sizes for meta-analysis. In H. Cooper, L. V. Hedges, and J. C. Valentine (Eds), *Handbook of research synthesis and meta-analysis, 3rd edition*. New York: Russell Sage Foundation, pp. 208-243. https://doi.org/10.7758/9781610448864.14

Boldt, A., & Gilbert, S. J. (2019). Confidence guides spontaneous cognitive offloading. *Cognitive Research: Principles and Implications*, *4*(1), 45. https://doi.org/10.1186/s41235-019-0195-y

Bouriga, S., & Olive, T. (2021). Is typewriting more resources-demanding than handwriting in undergraduate students? *Reading and Writing*, 1-29. https://doi.org/10.1007/s11145-021-10137-6

Borenstein, M., Higgins, J. P., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Research synthesis methods, 8*(1), 5-18. https://doi.org/10.1002/jrsm.1230

Broe, D. (2013). The effect of teaching Cornell notes on student achievement [Masters thesis, Minot State University]. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1061.8854&rep=rep1&type=pdf

Brown, C. M. (1988). Comparison of typing and handwriting in "two-finger typists". *Proceedings of the Human Factors Society Annual Meeting 32*(5), 381-385. https://doi.org/10.1177/154193128803200533

*Bui, D. C., Myerson, J., & Hale, S. (2013). Note-taking with computers: Exploring alternative strategies for improved recall. *Journal of Educational Psychology*, *105*(2), 299-309. https://doi.org/10.1037/a0030367

*Chiaraluce, B. (2019). Effects of distraction on memory while handwriting or typing notes (Publication No. 2217838616) [Southern Connecticut State University]. ProQuest Dissertations & Theses Global.

Clark, W. (2008). *Academic charisma and the origins of the research university*. Chicago, IL: University of Chicago Press. https://doi.org/10.7208/9780226109237

Connelly, V., Dockrell, J. E., & Barnett, J. (2005). The slow handwriting of undergraduate students constrains overall performance in exam essays. *Educational Psychology, 25*(1), 99-107. https://doi.org/10.1080/0144341042000294912

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior, 11*(6), 671-684. https://doi.org/10.1016/S0022-5371(72)80001-X

Crawford, C. C. (1925). Some experimental studies of the results of college note-taking. *The Journal of Educational Research*, *12*(5), 379-386. https://doi.org/10.1080/00220671.1925.10879612

*Crumb, R. M., Hildebrandt, R., & Sutton, T. M. (2022). The value of handwritten notes: A failure to find state-dependent effects when using a laptop to take notes and complete a quiz. *Teaching of Psychology, 49*(1), 7-13. https://doi.org/10.1177/0098628320979895

*Cubilo, J. (2017). Video-mediated listening passages and typed note-taking: Examining their effects on examinee listening test performance and item characteristics (Publication No. 2014017441)[University of Hawaii at Manoa]. ProQuest Dissertations & Theses Global.

Cunningham, A. E., & Stanovich, K. E. (1990). Assessing print exposure and orthographic processing skill in children: A quick measure of reading experience. *Journal of Educational Psychology, 82*(4), 733. https://doi.org/10.1037/0022-0663.82.4.733

*Desselle, S. P., & Shane, P. A. (2018). Laptop versus longhand note taking in a professional doctorate course: student performance, attitudes, and behaviors. *INNOVATIONS in Pharmacy, 9*(3), 1. https://doi.org/10.24926/iip.v9i3.1392

*Devers, C., Hoffman, J., Lee, C., Ragsdale, E., & Devers, E. (2015, October). Note-Taking: Paper Versus Computer. *In E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (pp. 772-776). Association for the Advancement of Computing in Education (AACE). http://www.learntechlib.org/p/152092/

Dhakal, V., Feit, A. M., Kristensson, P. O., & Oulasvirta, A. (2018). Observations on typing from 136 million keystrokes. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems,* 1-12. https://doi.org/10.1145/3173574.3174220

Dinsmore, D. L., Alexander, P. A. (2012). A Critical Discussion of Deep and Surface Processing: What It Means, How It Is Measured, the Role of Context, and Model Specification. *Educational Psychology Review, 24*, 499–567. https://doi.org/10.1007/s10648-012-9198-7

Di Vesta, F. J., & Gray, G. S. (1972). Listening and note taking. *Journal of Educational Psychology, 63*(1), 8-14. https://doi.org/10.1037/h0032243

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. *Psychological Science in the Public Interest, 14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Van Drempt, N., McCluskey, A., & Lannin, N. A. (2011). A review of factors that influence adult handwriting performance. *Australian Occupational Therapy Journal, 58*(5), 321-328. https://doi.org/10.1111/j.1440-1630.2011.00960.x

Dror, I. E. (2007). Introduction Gold mines and land mines in cognitive technology. In I. E. Dror (Ed.), *Cognitive Technologies and the Pragmatics of Cognition* (pp. 1-7). John Benjamins Publishing Company. https://doi.org/10.1075/bct.12.02dro

*Duran, K. S., & Frederick, C. M. (2013). Information comprehension: Handwritten vs. typed notes. *Undergraduate Research Journal for the Human Sciences, 12*(1). https://www.kon.org/urc/v12/duran.html

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455-463. https://doi.org/10.1080/01621459.2000.10473905

*Eason, T. (2017). *Various methods of note-taking and how they compare in terms of information retention* (Publication No. 1952349063)[Undergraduate Thesis, East Carolina State] ProQuest Dissertations & Theses Global.

Eisner, S., & Rohde, K. (1959). Note taking during or after the lecture. *Journal of Educational Psychology, 50*(6), 301-304. https://doi.org/10.1037/h0038628

Fiorella, L., & Mayer, R. E. (2017). Spontaneous spatial strategy use in learning from scientific text. *Contemporary Educational Psychology*, *49*, 66-79. https://doi.org/10.1016/j.cedpsych.2017.01.002

Fisher, J. L., & Harris, M. B. (1973). Effect of note taking and review on recall. *Journal of Educational Psychology, 65*(3), 321–325. https://doi.org/10.1037/h0035640

*Flanigan, A. E., & Titsworth, S. (2020). The impact of digital distraction on lecture note taking and student learning. *Instructional Science, 48*(5), 495-524. https://doi.org/10.1007/s11251-020-09517-2

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34*(10), 906-911. https://doi.org/10.1037/0003-066X.34.10.906

Gong, C., & Rodd, J. M. (2020). *A cognitive offloading perspective on the modality effect in note-taking.* PsyArxiv. https://doi.org/10.31234/osf.io/qwdpf

Greene, E. B. (1928). The relative effectiveness of lecture and individual reading as methods of college teaching. *Genetic Psychology Monographs - Child Behavior and Genetic Psychology, 4*(6), 459-523.

Greene, E. B. (1934). Certain aspects of lecture, reading and guided reading. *School and Society, 39*, 619–624.

Gross, T. (2014, December 30). This year, I resolve to ban laptops from my classroom. *The Washington Post.* https://www.washingtonpost.com/posteverything/wp/2014/12/30/this-year-im-resolving-to-ban-laptops-from-my-classroom/

*Gür, T. (2021). The Effect of Verbatim and Generative Notes Taken by Hand and Keyboard at University Level on Success and Persistence. *Education Quarterly Reviews, 4*(3). https://doi.org/10.31219/osf.io/dw5q7

Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological methods, 3*(4), 486. https://doi.org/10.1037/1082-989X.3.4.486

Henk, W. A., & Stahl, N. A. (1985). A meta-analysis of the effect of notetaking on learning from lectures. *National Reading Conference Yearbook, 34*, 70–75. https://eric.ed.gov/?id=ED258533

Henmi, M., & Copas, J. B. (2010). Confidence intervals for random effects meta-analysis and robustness to publication bias. *Statistics in Medicine, 29*(29), 2969-2983. https://doi.org/10.1002/sim.4029

Hartley, J., & Davies, I. K. (1978). Note-taking: A critical review. *Programmed Learning and Educational Technology, 15*(3), 207-224. https://doi.org/10.1080/0033039780150305

Harville, D. (1977). Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *Journal of the American Statistical Association*, *72*(358), 320-338. https://doi.org/10.1080/01621459.1977.10480998

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction*, *7*(2), 174-196. https://doi.org/10.1145/353485.353487

Horbury, S. R., & Edmonds, C. J. (2021). Taking class notes by hand compared to typing: effects on children's recall and understanding. *Journal of Research in Childhood Education, 35*(1), 55-67. https://doi.org/10.1080/02568543.2020.1781307

Ioannidis, J. P., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical Trials, 4*(3), 245-253. https://doi.org/10.1177/1740774507079441

James, K. H., & Atwood, T. P. (2009). The role of sensorimotor learning in the perception of letter-like forms: Tracking the causes of neural specialization for letters. *Cognitive Neuropsychology, 26*(1), 91-110. https://doi.org/10.1080/02643290802425914

*Jamet, E., Gonthier, C., Cojean, S., Colliot, T., & Erhel, S. (2020). Does multitasking in the classroom affect learning outcomes? A naturalistic study. *Computers in Human behavior, 106*, 106264. https://doi.org/10.1016/j.chb.2020.106264

Jay, T., Caldwell-Harris, C., & King, K. (2008). Recalling taboo and nontaboo words. *The American journal of psychology, 121*(1), 83-103. https://doi.org/10.2307/20445445

Jones, H. E. (1923). *Experimental studies of college teaching - The effect of examination on permanence of learning* (Publication No. 301801226) [Doctoral dissertation, Columbia University]. ProQuest Dissertations & Theses Global.

Jones, E. S. (1927). Testing and training the inferior freshman. *Personnel Journal, 6*(1), 182-191. http://www.worldcat.org/oclc/2960927

Karat, C. M., Halverson, C., Horn, D., Karat, J. (1999). Patterns of entry and correction in large vocabulary continuous speech recognition systems. *Proceedings of the 1999 SIGCHI conference on Human Factors in Computing Systems*, 568-575. https://doi.org/10.1145/302979.303160

*Kennedy, D. R. (2019). The effect of banning computers on examination performance in a first-year pathophysiology class. *American Journal of Pharmaceutical Education, 83*(10). https://doi.org/10.5688/ajpe7535

*Kirkland, K. M. (2016). The effect of note taking media and preference on the cognitive processes involved in learning. [Undergraduate thesis, University of Colorado, Boulder]. https://scholar.colorado.edu/concern/undergraduate_honors_theses/6t053g62z

Kiewra, K. A. (1989). A review of note-taking: The encoding-storage paradigm and beyond. *Educational Psychology Review, 1*(2), 147-172. https://doi.org/10.1007/BF01326640

Kiewra, K. A., DuBois, N. F., Christian, D., McShane, A., Meyerhoffer, M., & Roskelley, D. (1991). Note-taking functions and techniques. *Journal of Educational Psychology, 83*(2), 240. https://doi.org/10.1037/0022-0663.83.2.240

*Kodaira, Y. (2017). An investigation of cognitive processes associated with notetaking and notes-review [Doctoral Dissertation, Columbia University]. https://doi.org/10.7916/D85M6J31

Kobayashi, K. (2006). Combined effects of note-taking/-reviewing on learning and the enhancement through interventions: A meta-analytic review. *Educational Psychology, 26*(3), 459-477. https://doi.org/10.1080/01443410500342070

Kossmeier M, Tran U, Voracek M (2020). *metaviz: Forest Plots, Funnel Plots, and Visual Funnel Plot Inference for Meta-Analysis*. R package version 0.3.1, https://CRAN.R-project.org/package=metaviz.

*Kuipers, I. (2019). *Aantekeningen maken bij een college: het effect van medium, gender en kwaliteit* [Undergraduate Thesis, Tilburg University]. TU campus repository. http://arno.uvt.nl/show.cgi?fid=150460

*Kutta, T. J. (2017). *Note-taking and technology* (Publication No. 1984971954)[Doctoral Thesis, Florida State University]. ProQuest Dissertations & Theses Global.

*Lalchandani, L. A. (2019). *Elucidating the cognitive processes involved in the note-taking effect* (Publication No. 2232683536)[Doctoral Dissertation, University of Colorado at Boulder]. ProQuest Dissertations & Theses Global.

Longcamp, M., Zerbato-Poudou, M. T., & Velay, J. L. (2005). The influence of writing practice on letter recognition in preschool children: A comparison between handwriting and typing. *Acta Psychologica, 119*(1), 67-79. https://doi.org/10.1016/j.actpsy.2004.10.019

van der Loo M (2014). The stringdist package for approximate string matching. *The R Journal, 6*, 111-122. https://journal.r-project.org/archive/2014/RJ-2014-011

*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*(6), 947-971. https://doi.org/10.1007/s11251-018-9458-0

MacKenzie, I. S., & Soukoreff, R. W. (2002). Text Entry for Mobile Computing: Models and Methods,Theory and Practice. *Human-Computer Interaction*, *17*(2-3), 147-198. http://doi.org/10.1207/S15327051HCI172&3_2

MacKenzie, I. S., & Zhang, S. X. (1999). The design and evaluation of a high-performance soft keyboard. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 25-31). https://doi.org/10.1145/302979.302983

Makany, T., Kemp, J., & Dror, I. E. (2009). Optimising the use of note-taking as an external cognitive aid for increasing learning. *British Journal of Educational Technology, 40*(4), 619-635. https://doi.org/10.1111/j.1467-8535.2008.00906.x

May, C. (2014, June 3). A learning secret: Don't take notes with a laptop. *Scientific American.* https://www.scientificamerican.com/article/a-learning-secret-don-t-take-notes-with-a-laptop/

McClendon, P. I. (1956). *An experimental study of the relationship between the note-taking practices and listening comprehension of college freshmen during expository lectures* (Publication No. 302001903) [Doctoral dissertation, University of Iowa]. ProQuest Dissertations & Theses Global. https://doi.org/10.1080/03637755809375236

Meyer, R. (2014, May 1). To remember a lecture better, take notes by hand. *The Atlantic.* https://www.theatlantic.com/technology/archive/2014/05/to-remember-a-lecture-better-take-notes-by-hand/361478/

*Mitchell, A., & Zheng, L. (2019). Examining longhand vs. laptop debate: A replication study. *AIS Transactions on Replication Research, 5*(9). https://doi.org/10.17705/1atrr.00041

*Morehead, K., Dunlosky, J., & Rawson, K. A. (2019). 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*(3), 753-780. https://doi.org/10.1007/s10648-019-09468-2

*Mueller, P. A., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. *Psychological Science*, *25*(6), 1159-1168. https://doi.org/10.1177/0956797614524581

*Murphy, D. (2016, April 8-12). Handwriting and Typing: The Impact of Note-Taking on Recall [Conference paper]. AERA, Washington, DC, United States. https://eric.ed.gov/?id=ED592169

Paek, J. (2008). *Design of soft keyboards for reducing visual attention shifts* [Masters thesis, Korea Advanced Institute of Science and Technology]. KAIST Open Access Self-Archiving System. http://hdl.handle.net/10203/54949

R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Re, A. C. D. (2013). *compute.es: Compute Effect Sizes*. https://cran.r-project.org/package=compute.es

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, *20*(9), 676-688. https://doi.org/10.1016/j.tics.2016.07.002

*Schoen, I. (2012). *Effects of method and context of note-taking on memory: Handwriting versus typing in lecture and textbook-reading contexts* [Undergraduate Thesis, Pitzer College]. https://scholarship.claremont.edu/pitzer_theses/20/

Schultz, C. B., & Di Vesta, F. J. (1972). Effects of passage organization and note taking on the selection of clustering strategies and on recall of textual materials. *Journal of Educational Psychology, 63*(3), 244. https://doi.org/10.1037/h0032651

*Sheppard, A. (2015). Note-taking strategies in a lecture environment []. http://archives.algomau.ca/main/sites/default/files/2015-035_003_013_001.pdf

Seki, K., Yajima, M., & Sugishita, M. (1995). The efficacy of kinesthetic reading treatment for pure alexia. *Neuropsychologia, 33*(5), 595-609. https://doi.org/10.1016/0028-3932(94)00138-F

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*(6), 592-604. https://doi.org/10.1037/0278-7393.4.6.592

*Slavina, A. (2018). Note-taking and the decision to externalize memory [Doctoral dissertation, Iowa State University]. https://dr.lib.iastate.edu/handle/20.500.12876/30853

Soukoreff, W. R., & Mackenzie, I. S. (1995). Theoretical upper and lower bounds on typing speed using a stylus and a soft keyboard. *Behaviour & Information Technology, 14*(6), 370-379. https://doi.org/10.1080/01449299508914656

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science, 333*(6043), 776-778. https://doi.org/10.1126/science.1207745

Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.) *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 99-110). Wiley. https://doi.org/10.1002/0470870168.ch6

Strauss, V. (2016, August 5). High school teacher: I'm banning laptops in class — and not just because they are distracting. *The Washington Post.* https://www.washingtonpost.com/news/answer-sheet/wp/2016/08/05/high-school-teacher-im-banning-laptops-in-class-and-not-just-because-they-are-distracting/

*Tischner, C. (2017). Note-taking method and recall performance: Effects of computer-based versus handwritten note-taking [Undergraduate Thesis, Northern Kentucky University]. https://dspace.nku.edu/bitstream/handle/11216/2458/Tischner%20HIP.pdf?sequence=1

Trochim, W. M. (1989). An introduction to concept mapping for planning and evaluation. *Evaluation and program planning, 12*(1), 1-16. https://doi.org/10.1016/0149-7189(89)90016-5

*Urry, H. L., Crittle, C. S., Floerke, V. A., Leonard, M. Z., Perry III, C. S., Akdilek, N., ... & Zarrow, J. E. (2021). Don't ditch the laptop just yet: A direct replication of Mueller and Oppenheimer's (2014) study 1 plus mini meta-analyses across similar studies. *Psychological Science, 32*(3), 326-339. https://doi.org/10.1177/0956797620965541

Valentine, J. C. (2019). Incorporating judgments about study quality into research syntheses. In H. Cooper, L.V. Hedges, and J. C. Valentine (Eds.), *The Handbook of Research Synthesis and Meta-Analysis, 3rd edition*, pp. 129-140. Russell Sage Foundation. https://doi.org/10.7758/9781610448864.10

Valentine, J. C., Aloe, A. M., & Lau, T. S. (2015). Life after NHST: How to describe your data without "p-ing" everywhere. *Basic and Applied Social Psychology*, *37*, 260-273. https://doi.org/10.1080/01973533.2015.1060240

Vernon, P. E. (1946). An experiment on the value of the film and film-strip in the instruction of adults. *The British Journal of Educational Psychology 16*(3) 149-162. https://doi.org/10.1111/j.2044-8279.1946.tb01598.x

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics, 30(3)*, 261–293. https://doi.org/10.3102/10769986030003261

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1-48. https://doi.org/10.18637/jss.v036.i03

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods, 1*(2), 112-125. https://doi.org/10.1002/jrsm.11

Wegner, D. M. (1986). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G.R. Goethals (Eds.), *Theories of Group Behavior*, (pp. 185-208). Springer Verlag. https://doi.org/10.1007/978-1-4612-4634-3_9

Wagner, M. E., & Strabel, E. (1935). Teaching high-school pupils how to study. *The School Review*, *43*(8), 577-589. https://doi.org/10.1086/439835

*Wei, F. Y. F., Wang, Y. K., & Fass, W. (2014). An experimental study of online chatting and notetaking techniques on college students' cognitive learning from a lecture. *Computers in Human Behavior, 34*, 148-156. https://doi.org/10.1016/j.chb.2014.01.019

What Works Clearinghouse (2022). What Works Clearinghouse Procedures and Standards Handbook, Version 5.0. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). This report is available on the What Works Clearinghouse website at https://ies.ed.gov/ncee/wwc/Handbooks.

Wickham H (2019). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.4.0, https://CRAN.R-project.org/package=stringr.

Wickham H, François R, Henry L, Müller K (2022). *dplyr: A Grammar of Data Manipulation*. R package version 1.0.9, https://CRAN.R-project.org/package=dplyr.

*Wood, E., Zivcakova, L., Gentile, P., Archer, K., De Pasquale, D., & Nosko, A. (2012). Examining the impact of off-task multi-tasking with technology on real-time classroom learning. *Computers & Education, 58*(1), 365-374. https://doi.org/10.1016/j.compedu.2011.08.029

TABLES

Table 1

Influences and constraints that might covary with modality assignment

Influence/constraint	Type	Description	Discussion
Source type	External	Nature of the source or sources of information, e.g., live lecture, audio, book, etc.	All sources require switching attention between the source itself and transcription, but some sources, e.g., lecture with slides, require switching attention between multiple sources and transcription.
Length	External	Amount of time information is being shared.	When information is shared for prolonged periods of time, the cognitive resources can be taxed and learning may be less effective.
Pace	External	Speed at which information is being shared.	When information is shared the speed at which it is distributed can vary and higher speeds are more taxing on cognitive resources.
Intermittence	External	Pauses in the sharing of information.	The frequency and length of pauses in speech affect the pace and might provide note-takers additional time to process and synthesize the information. Participants watching recorded material may have the ability to control intermittence (by pausing the recording).
Novelty/Familiarity	External	How familiar the note-taker is with the information being shared.	Note-takers may take qualitatively different notes and use different approaches if they are more or less familiar with the information being shared and in addition, familiarity leads to

Influence/constraint	Type	Description	Discussion
			more effective long-term storage and retrieval
Scaffold/Framework	External	Documents which guide or partially complete notes.	Partial notes may reduce the transcription burden on note-takers and allow them to focus more resources on annotation, processing, and encoding information being shared.
Primary language	External	Information source primary language compared with the note-taker's.	The primary language of the note-taker in reference to the language of the incoming information may play a role in additional cognitive demands that they have to allocate to note-taking.
Language complexity	External	Complexity of the information source.	The complexity of the language used for note-taking may play a role in additional cognitive resource demand.
Distraction	External	software / internet distractions from incoming information	There are a number of studies in the note-taking literature discussing the negative effects that distracted students on electronic devices have to cope with that handwriters don't have, in educational settings.
Transcription capacity	Internal	Speed of the note-taker to take notes.	Note-taking speed enables alternative note-taking strategies which can provide additional time for processing and synthesis.
Processing speed	Internal	Speed of the note-taker to process the incoming information.	Processing speed in conjunction with pace modulates note-taking style and other meta-cognitive decisions.
Working memory capacity	Internal	Amount of information able to be briefly stored.	Working memory capacity varies across participants and affects the amount of synchronous synthesis that participants are able to do and the quality of verbatim notes.
Short & Long term	Internal	Amount of information able to be	Short & Long term memory varies across participants and affects

Influence/constraint	Type	Description	Discussion
memory		temporarily stored.	the ability of the note-taker to maintain information that was received while thinking about other subjects and the ability to store information for later retrieval.
Prior experience with mediums	Internal	Experience with different note-taking mediums.	Note-takers' experience with different note-taking mediums can affect their concentration, cognitive load, and transcription capacity.
Intent	Internal	Reason for note-taking.	Note-takers' intention behind taking notes can affect the quality, quantity, and type of notes, as well as the metacognitive strategies behind what is transcribed.
Review	Internal	Reviewing notes that have been taken.	Note-review is a behavior often present during note-taking for the purposes of editing, proofreading, connecting concepts, ideas, and information, as well as for enhanced encoding benefits. It is also used as a strategy post-note-taking for further encoding benefits.

Table 2

Meta-table of study characteristic tables

Table number	Table title	Table content
2.1	Publications over time	Counts of distinct report identifiers, sample identifiers, measurement identifiers, and effect size identifiers over years.
2.2	Report characteristics, setting, sample	Year study was released, publication type (dissertation vs journal article vs other report), country conducted in, type of environment e.g., college/university vs secondary school vs primary school, average age or grade level, if US race/ethnicity (or study author description of the race/ethnicity of sample), % female.
2.3	Independent variable considerations, Status on potential moderators	Distractor, setting e.g., lab vs classroom, source audio only, visual only, audio + visual, if live, opportunities to participate, Transcription capacity, speaker pace, note-taking style, opportunity to review notes, and the note-takers' intention with their notes.
2.4	Research design and study quality, measurement	Assignment mechanism, attrition within and across groups, reliability (coefficient alpha, interrater reliability/agreement); variables used as statistical controls for effect size estimation, factual vs conceptual, multiple choice vs. short answer vs true/false vs. long answer, any reliability information, any validity information.
2.5	Effect size information	Effect size, standard error, sample sizes

Table 2.1

Publication count over time

Year	dist_rid	dist_sid	dist_mid	dist_eid
1987	1	1	1	2
2012	2	2	2	4
2013	2	3	5	5
2014	3	6	8	13
2015	2	2	2	2
2016	4	5	8	13
2017	5	6	7	12
2018	3	4	4	5
2019	5	6	7	17
2020	4	4	7	9
2021	2	3	4	6

Notes.

dist_rid = distinct count of report identification numbers.

dist_sid = distinct count of sample identification numbers.

dist_mid = distinct count of measure identification numbers.

dist_eid = distinct count of effect size identification numbers.

Table 2.2

Sample characteristics

Author(s)	Publication Type	Country	Environment	Setting	Age (mean)	Grade	Race	% Female
Artz et al., 2020	Journal	USA	University	Classroom	25	Undergraduates	[l]	36%
Bargione, 1987	Dissertation	USA	Secondary	Lab	16	High School	[m]	100%
Beck, 2014	Journal	USA	University	Classroom	25	Undergraduates	---	76%
Blankenship, 2016	Dissertation	USA	University	Classroom	---	Undergraduates	[n]	85%
Bui et al., 2013	Journal	USA	University	Lab	19	Undergraduates	---	66%
Chiaraluce, 2019	Thesis	USA	University	Lab	19	Undergraduate	---	66%
Crumb et al., 2020	Journal	USA	University	Lab	19	Undergraduates	---	55%
Cubillo, 2017	Dissertation	USA	University	Lab	22	Undergraduates	[o]	54%
Desselle & Shane, 2018	Journal	USA	University	Classroom	---	Undergraduates	[p]	64%
Devers, 2015	Other: Conference	USA	University	Lab	---	[g]	---	78%
Duran & Frederick, 2013	Journal	USA	University	Lab	[a]	Undergraduates	---	---
Eason, 2017	Thesis	USA	University	Lab	19	Undergraduates	[q]	62%
Flanagan & Titsworth, 2020	Journal	USA	University	Lab	---	Undergraduates	[r]	61%
Gür, 2021	Journal	Turkey	University	---	---	---	---	48%
Jamet et al., 2020	Journal	France	University	Classroom	19	Undergraduates	---	83%
Kennedy, 2019	Journal	USA	University	Classroom	---	[h]	---	63%
Kirkland, 2016	Thesis	USA	University	Classroom	---	Undergraduates	---	---

Author(s)	Publication Type	Country	Environment	Setting	Age (mean)	Grade	Race	% Female
Kodaira, 2017	Dissertation	USA	University	Lab	19	[i]	[s]	85%
Kuipers, 2019	Thesis	Netherlands	Secondary	Classroom	17	Secondary	---	51%
Kutta, 2017	Dissertation	USA	University	Lab	---	Undergraduates	---	---
Lalchandani, 2016	Thesis	USA	University	Lab	---	Undergraduates	---	---
Luo et al., 2018	Journal	Germany	University	Classroom	---	[j]	[t]	80%
Mitchell & Zheng, 2019	Journal	USA	University	Classroom	---	Undergraduates	---	51%
[1]Morehead et al., 2019	Journal	USA	University	Lab	[b]	Undergraduates	---	76%
[2]Morehead et al., 2019 exp 2	Journal	USA	University	Lab	[c]	Undergraduates	---	84%
[1]Mueller & Oppenheimer, 2014	Journal	USA	University	Lab	---	[k]	---	50%
[2]Mueller & Oppenheimer, 2014	Journal	USA	University	Lab	---	[k]	---	77%
[3]Mueller & Oppenheimer, 2014	Journal	USA	University	Lab	---	[k]	---	75%
Murphy, 2016	Other: Conference	USA	Secondary	Classroom	[d]	HS Seniors	---	43%
Schoen, 2012	Thesis	USA	University	Lab	[e]	Undergraduates	---	76%
Sheppard, 2015	Thesis	Canada	University	Lab	20	Undergraduates	---	73%
Slavina, 2018	Dissertation	USA	University	Lab	19	Undergraduates	---	60%
Tischner, 2017	Thesis	USA	Other: Remote	Other: Remote	---	Undergraduates	---	---
Urry et al., 2021	Journal	USA	University	Lab	---	Undergraduates	[u]	62%
Wei et al., 2014	Journal	USA	University	Lab	22	Undergraduates	[v]	53%

Author(s)	Publication Type	Country	Environment	Setting	Age (mean)	Grade	Race	% Female
Wood et al., 2012	Journal	Canada	University	Classroom	[f]	Undergraduates	---	80%

Notes. Superscript numbers or letters in the sample descriptor cells provide detailed information on sample characteristics.

[1] Experiment 1

[2] Experiment 2

[3] Experiment 3

[a] "aged 18-26"

[b] "...ages ranged from 18 to 44 (84% of participants were under the age of 22)." p.6

[c] "...ages ranged from 18 to 38 with 90% of participants under the age of 22." p.14

[d] "High school seniors from one northeast area Baltimore County high school, who were at least eighteen years old..." p.5

[e] "ages 18-23"

[f] "...males (Mage=20.67, SD=2.33) and females (Mage=19.56, SD=1.19)"

[g] "There were 24 first-year, 8 second-year, 8 third-year, 5 fourth-year, and 13 graduate students." p. 774

[h] "1st year pharmacy students"

[i] "The majority reported being in their first year of school (57.5%), followed by students in their second year (30.0%), with the remainder in their third (8.75%) or fourth year (1.25%), or beyond (2.5%)." p. 37

[j] "Seventy-one percent were juniors and seniors..."

[k] "university subject pool participants"

[l] "Nonwhite: =1 if student's race is not white and 0 otherwise 0.105 (0.308)" p. 107

[m] "There were 24 whites (60%) and 16 (40%) blacks."

[n] "...participants were primarily female (85%) and Caucasian (90%)." p.38

[o] "The examinees came from 22 different first language backgrounds including Chinese (40.5%), Korean (20%), Japanese (14.5%), Spanish (5.5%), Arabic (4.5%), Vietnamese (2.5%), Turkish (2.5%), Cantonese (1.5%), German (1.5%)..." p.62

[p] "Race/ethnicity Asian/Pacific Islander 55 (63.4) Black 5 (5.8) White 19 (22.1) Hispanic 4 (4.7) No Response/unknown 3 (3.5)" p.9

[q] "90.1% of the participants were not of Hispanic/Latino ethnicity, and 9.9% of the population was Hispanic/Latino. Racially, the participants identified as follows: 75.3% white, 11.1% Black, 11.1% Asian, and 2.4% other races." p.11

[r] "Most participants were white (86%)" p.4

[s] "...the racial/ethnic distribution was 81.25% White, 10.0% Asian, 3.75% Black/African American, 1.25% Hispanic, and 3.75% Other." p. 37

[t] "...most were Caucasians (94%)."

[u] "...5% were African American or Black, 24% were Asian, 58% were White, 5% were Hispanic or Latinx, and 7% were multiracial; two people declined to report their race/ethnicity." p. 328

[v] "79.6% were Caucasian, 12.6% were African American, and the remaining were of other ethnicities." p.151

Table 2.3

Study characteristics

Author(s)	Setting	Source	Recording	Live Lecture	Distractor	Style	Transcription Capacity	Speaker Pace	Review	Intention
Artz et al., 2020	Class	*A + V	No	True	No	---	---	No	Yes	Yes
Bargione, 1987	Lab	A + V	Yes	False	No	---	d	No	No	---
Beck, 2014	Class	A + V	Yes	False	No	---	---	No	---	---
Blankenship, 2016	Class	A + V	Yes	False	Yes	---	---	No	No	---
Bui et al., 2013	Lab	AUDIO	Yes	False	No	Summary	---	No	No	---
Bui et al., 2013	Lab	AUDIO	Yes	False	No	Verbatim	---	No	No	---

Author(s)	Setting	Source	Recording	Live Lecture	Distractor	Style	Transcription Capacity	Speaker Pace	Review	Intention
Chiaraluce, 2019	Lab	A + V	Yes	False	Yes	---	---	No	No	---
Crumb et al., 2020	Lab	A + V	Yes	False	No	---	---	No	Yes	---
Cubillo, 2017	Lab	A + V	Yes	False	No	---	---	No	Yes	---
Desselle & Shane, 2018	Class	A + V	No	True	No	---	---	No	Yes	Yes
Devers, 2015	Lab	A + V	Yes	False	No	---	---	No	No	---
Duran & Frederick, 2013	Lab	A + V	Yes	False	No	---	---	No	---	---
Eason, 2017	Lab	A + V	Yes	False	[a]	---	---	No	No	---
Flanagan & Titsworth, 2020	Lab	A + V	Yes	False	Yes	---	---	No	Yes	---
Gür, 2021	---	---	---	---	---	Verbatim	---	---	---	---
Gür, 2021	---	---	---	---	---	Summary	---	---	---	---
Jamet et al., 2020	Class	Other	Other	False	[b]	---	---	No	No	---
Kennedy, 2019	Class	A + V	No	True	No	---	---	No	Yes	Yes
Kirkland, 2016	Class	A + V	Yes	False	Yes	---	---	No	---	---

Author(s)	Setting	Source	Recording	Live Lecture	Distractor	Style	Transcription Capacity	Speaker Pace	Review	Intention
Kodaira, 2017	Lab	A + V	Yes	False	No	---	[e]	No	Yes	Yes
Kuipers, 2019	Class	A + V	Yes	False	Yes	---	[f]	No	Yes	---
[1]Kutta, 2017	Lab	A + V	Yes	False	No	---	---	No	No	---
[2]Kutta, 2017	Lab	A + V	Yes	False	No	---	---	No	Yes	---
[1]Lalchandani, 2016	Lab	AUDIO	Yes	False	Yes	---	---	No	No	---
[2]Lalchandani, 2016	Lab	A + V	Yes	False	Yes	---	---	No	No	---
[1]Luo et al., 2018	Class	A + V	Yes	False	Yes	---	---	No	Yes	Yes
[2]Luo et al., 2018	Class	A + V	Yes	False	Yes	---	---	No	No	No
Mitchell & Zheng, 2019	Class	A + V	Yes	False	[c]	---	---	---	No	---
Morehead et al., 2019	Lab	A + V	Yes	False	Yes	---	[g]	No	---	No
[1]Mueller & Oppenheimer, 2014	Lab	A + V	Yes	False	Yes	---	---	No	No	---
[2]Mueller & Oppenheimer, 2014	Lab	A + V	Yes	False	Yes	---	[h]	No	No	---
[2]Mueller & Oppenheimer, 2014	Lab	A + V	Yes	False	Yes	Summary	[h]	No	No	---

Author(s)	Setting	Source	Recording	Live Lecture	Distractor	Style	Transcription Capacity	Speaker Pace	Review	Intention
[3]Mueller & Oppenheimer, 2014	Lab	A + V	Yes	False	No	---	---	No	Yes	---
[2]Mueller & Oppenheimer, 2014	Lab	A + V	Yes	False	No	---	---	No	No	---
Murphy, 2016	Class	A + V	Yes	False	Yes	---	---	No	No	---
Schoen, 2012	Lab	A + V	Yes	False	Yes	---	[i]	No	No	---
Sheppard, 2015	Lab	A + V	Yes	False	No	---	---	No	---	---
Slavina, 2018	Lab	A + V	Yes	False	No	---	---	No	No	Yes
Tischner, 2017	Other	A + V	Yes	False	Yes	---	---	No	Yes	---
Urry et al., 2021	Lab	A + V	Yes	False	Yes	---	---	No	---	---
Wei et al., 2014	Lab	A + V	Yes	False	No	---	---	No	---	---
Wood et al., 2012	Class	A + V	No	True	No	---	---	No	---	---

Notes. Superscript symbols, numbers, and letters in the sample descriptor cells provide detailed information on sample characteristics.

[1] Experiment 1

[2] Experiment 2

[3] Experiment 3

* A + V = Audio plus visual input for participants

[a] "Once the presentation was over, the participants were given access to an online questionnaire that gathered demographic information, asking whether or not they had seen the presentation before, and also to assess their NCog rating using the Need for Cognition Scale (Cacioppo & Petty, 1982) and optimal arousal levels using the BMIS (Mayer & Gaschke, 1988)." p.16

[b] "maybe; filled out 3 parts of questionnaire prior to learning outcomes section"

[c] "After watching the video, students participated in regular classroom activities (i.e., lecture, notes, and in class activities) related to the regular content of the course (i.e., IS or Economics). The topic of the video was not discussed concluding the viewing. Approximately 30 minutes after watching the video, students took a closed-note quiz..." p. 4

[d] "To control for typing ability each subject was given a typing test prior to the experiment. Subjects were placed into one of two subgroups (i.e., 30-40 words per minute or 50+ words per minute). Twenty subjects with high typing skill and 20 subjects with low typing skill were assigned randomly to two experimental groups: (a) note taking by microcomputer (10 high skill and 10 low skill typists), and (b) note taking by hand (10 high skill and 10 low skill typists)."

[e] "For subjects taking notes using a laptop, letter speed was measured using a modification of the Olinghouse and Graham (2009) task, whereby subjects were asked to access a word processing document to type..." p. 42

[f] "The typing test was taken on site where students had to write as many words as possible in a minute typing... Students in the written condition were asked to do as much as possible in one minute letter pairs, consisting of a capital letter and a small letter, to be noted below each other on order of the alphabet in their normal handwriting. Provide the legible and correct letter pairs up a point. The more points a student gets, the faster he can write. The writing test is based on Reddington et al. (2015)." pp.16-17

[g] "For the typing speed measure, participants were shown a passage and given 1 min to type as much of the passage as they could, and the number of words typed in 1 min was calculated." p.7

[h] "Participants then completed a typing test, the Need for Cognition scale..." p. 4

[i] "All participants reported dexterity in both handwriting notes and typing notes..."

Table 2.4

Research design and study quality

Author(s)	Reliability	Assignment	Question Measure	Question Type
[1]Mueller & Oppenheimer, 2014	irr = 0.89	Random	Factual	Other: Short Answer & Unclear
[1]Mueller & Oppenheimer, 2014	irr = 0.89	Random	Conceptual	Other: Short Answer & Unclear
Mueller & Oppenheimer, 2014	---	Random	Factual	Other: Short Answer & Unclear
Mueller & Oppenheimer, 2014	---	Random	Conceptual	Other: Short Answer & Unclear
Kodaira, 2017	---	Random	Factual	Multiple Choice
Kodaira, 2017	---	Random	Conceptual	Multiple Choice
Bui, Myers, & Hale, 2013	irr = 0.82	Random	Factual	Other: Free Recall
Bui, Myers, & Hale, 2013	---	Random	Factual	Short Answer
Chiaraluce, 2019	ca = 0.44	Random	Factual	Multiple Choice
Urry et al., 2021	irr = 0.98	Random	Factual	Short Answer
Urry et al., 2021	irr = 0.99	Random	Conceptual	Short Answer
Desselle & Shane, 2018	---	Not-Random	---	Multiple Choice
Sheppard, 2015	---	Random	Factual	Other: Mix Mc/Short Answer

Author(s)	Reliability	Assignment	Question Measure	Question Type
Morehead et al., 2019	---	Random	Factual	Short Answer
Morehead et al., 2019	---	Random	Conceptual	Short Answer
Bargione, 1987	---	Random	Factual	Multiple Choice
Schoen, 2012	---	Random	Factual	Multiple Choice
Wood et al., 2012	---	Random	Factual	Multiple Choice
Duran & Frederick, 2013	---	Not-Random	Factual	Multiple Choice
Beck, 2014	---	Random	Factual	Other: Mix Tf / Mc
Wei et al., 2014	---	Random	Factual	Multiple Choice
Blankenship, 2016	---	Random	[a]	Multiple Choice
Blankenship, 2016	---	Random	[a]	Short Answer
Kirkland, 2016	---	Not-Random	[a]	Multiple Choice
Cubillo, 2017	ca = 0.85	Random	[a]	Multiple Choice
Eason, 2017	---	Random	Factual	Multiple Choice
Kutta, 2017	---	Random	[a]	Short Answer
Kennedy, 2019	---	Not-Random	---	Other: Mix Mc/Short Answer

Author(s)	Reliability	Assignment	Question Measure	Question Type
Lalchandani, 2016	---	Not-Random	Factual	Multiple Choice
Lalchandani, 2016	---	Not-Random	Conceptual	Multiple Choice
Crumb et al., 2020	---	Random	Factual	Short Answer
Crumb et al., 2020	---	Random	Conceptual	Short Answer
Mitchell & Zheng, 2019	---	Not-Random	Factual	Other: Mix Mc/Short Answer
Mitchell & Zheng, 2019	---	Not-Random	Conceptual	Short Answer
Flanagan & Titsworth, 2020	ca = 0.779	Random	[a]	Multiple Choice
Jamet et al., 2020	---	Not-Random	Factual	Short Answer
Jamet et al., 2020	---	Not-Random	Conceptual	Short Answer
Luo et al., 2018	---	Random	[b]	Multiple Choice
Slavina, 2018	---	---	Factual	Short Answer
Artz et al., 2020	---	Random	[a]	Multiple Choice
Gür, 2021	---	Random	[c]	Multiple Choice
Murphy, 2016	---	Not-Random	---	---
Tischner, 2017	---	Random	Factual	Multiple Choice

Author(s)	Reliability	Assignment	Question Measure	Question Type
Tischner, 2017	---	Random	Factual	Short Answer
Tischner, 2017	---	Random	Factual	Other: Fill In The Blank
Devers, 2015	---	Random	[a]	Other: Mix Mc/Short Answer
Kuipers, 2019	---	---	Factual	Multiple Choice

Notes.

irr = interrater reliability

ca = cronbach's alpha

[1] experiment 1

[a] Other: mix Factual & Conceptual

[b] "...contained a mixture of fact, relation-ship, concept, and skill items." p.953

[c] = "The chance factor was tried to be minimized by using the fill-in-the-gap questions in all tests..." p. 136

Table 2.5

Effect size information

Author(s)	*g*	*SE*	*n - HW*	*n - TW*
Mueller & Oppenheimer, 2014	0.055	0.244	33.5	33.5
Mueller & Oppenheimer, 2014	0.340	0.246	33.5	33.5
Mueller & Oppenheimer, 2014	0.115	0.200	50.33	50.33
Mueller & Oppenheimer, 2014	-0.037	0.199	50.33	50.33
Mueller & Oppenheimer, 2014	0.406	0.201	50.33	50.33
Mueller & Oppenheimer, 2014	0.326	0.201	50.33	50.33
Mueller & Oppenheimer, 2014	0.707	0.279	27.25	27.25
Mueller & Oppenheimer, 2014	0.653	0.278	27.25	27.25
Mueller & Oppenheimer, 2014	0.033	0.271	27.25	27.25

Mueller & Oppenheimer, 2014	-0.161	0.271	27.25	27.25
Kodaira, 2017	0.496	0.227	40	40
Kodaira, 2017	0.321	0.225	40	40
Bui et al., 2013	0.000	0.316	20	20
Bui et al., 2013	-0.151	0.317	20	20
Bui et al., 2013	-1.240	0.345	20	20
Bui et al., 2013	-1.299	0.348	20	20
Chiaraluce, 2019	-0.044	0.277	30	23
Desselle & Shane, 2018	0.689	0.327	11	75
Desselle & Shane, 2018	0.583	0.326	11	75
Bargione, 1987	0.286	0.318	20	20

Bargione, 1987	0.272	0.318	20	20
Schoen, 2012	-1.350	0.380	17	17
Beck, 2014	0.136	0.437	10.5	10.5
Beck, 2014	0.121	0.437	10.5	10.5
Wei et al., 2014	0.624	0.350	20	15
Blankenship, 2016	0.155	0.221	52	34
Blankenship, 2016	0.457	0.223	52	34
Cubillo, 2017	-0.195	0.142	100	100
Crumb et al., 2020	0.110	0.224	40	40
Crumb et al., 2020	0.745	0.231	40	40
Mitchell & Zheng, 2019	0.365	0.117	150	145

Mitchell & Zheng, 2019	0.105	0.117	150	145
Flanagan & Titsworth, 2020	0.086	0.283	25	25
Jamet et al., 2020	0.319	0.147	90	97
Jamet et al., 2020	0.020	0.146	90	97
Artz et al., 2020	0.125	0.133	103	127
Artz et al., 2020	0.059	0.133	127	103
Artz et al., 2020	0.039	0.133	103	127
Artz et al., 2020	0.013	0.133	127	103
Gür, 2021	0.929	0.276	29	29
Gür, 2021	0.502	0.267	29	29
Gür, 2021	0.937	0.277	29	29

Gür, 2021	0.969	0.278	29	29
Murphy, 2016	0.305	0.279	30	23
Murphy, 2016	0.460	0.281	30	23
Kuipers, 2019	0.373	0.166	76	72
Urry et al., 2021	-0.026	0.211	74	68
Urry et al., 2021	-0.121	0.228	74	68
Sheppard, 2015	-0.228	0.394	20	20
Morehead et al., 2019	0.229	0.330	32	31
Morehead et al., 2019	0.106	0.393	32	31
Morehead et al., 2019	0.231	0.347	32	31
Morehead et al., 2019	-0.019	0.386	32	31

Morehead et al., 2019	0.158	0.399	31	33
Morehead et al., 2019	0.201	0.421	31	33
Morehead et al., 2019	0.143	0.322	31	31
Morehead et al., 2019	-0.258	0.350	31	31
Morehead et al., 2019	0.111	0.313	31	31
Morehead et al., 2019	-0.097	0.335	31	31
Morehead et al., 2019	-0.231	0.323	30	30
Morehead et al., 2019	-0.234	0.340	30	30
Wood et al., 2012	0.125	0.380	21	21
Wood et al., 2012	-0.032	0.431	21	21
Wood et al., 2012	-0.098	0.377	21	21

Duran & Frederick, 2013	0.299	0.346	36	36
Kirkland, 2016	-0.074	0.268	52.5	52.5
Eason, 2017	0.137	0.344	28	29
Eason, 2017	0.227	0.345	26	26
Kutta, 2017	0.031	0.307	33.33	33.33
Kutta, 2017	-0.034	0.321	33.33	33.33
Kutta, 2017	0.013	0.348	33.33	33.33
Kutta, 2017	-0.045	0.320	33.33	33.33
Kennedy, 2019	0.108	0.261	60	71
Lalchandani, 2016	-0.242	0.583	12	12
Lalchandani, 2016	0.162	0.564	12	12

Lalchandani, 2016	0.306	0.522	12	12
Lalchandani, 2016	-0.208	0.505	12	12
Lalchandani, 2016	0.298	0.610	12	12
Lalchandani, 2016	0.458	0.575	12	12
Lalchandani, 2016	0.129	0.560	12	12
Lalchandani, 2016	0.000	0.505	12	12
Luo et al., 2018	0.287	0.320	32	34
Luo et al., 2018	-0.181	0.337	30	30
Slavina, 2018	0.147	0.267	40	44
Tischner, 2017	0.058	0.399	17	32
Tischner, 2017	0.171	0.367	17	32

Tischner, 2017	0.350	0.512	17	32
Devers, 2015	0.418	0.262	30	28

Notes. HW = handwriting; TW = typewriting

Table 3

Platform and Databases Used in Literature Search

Platform	Databases
EBSCO	Academic Search Complete, Education Full Text (H.W. Wilson), ERIC, OpenDissertations, Psychology and Behavioral Sciences Collection, APA PsycINFO
Proquest	ABI/INFORM Collection (1971 - current), Alt-PressWatch (1970 - current), American Periodicals (1740 - 1940), APA PsycArticles® (1894 - current), ARTbibliographies Modern (ABM) (1974 - current), Business Market Research Collection (1986 - current), Career & Technical Education Database, Coronavirus Research Database, Dissertations & Theses @ University of Louisville, Early Modern Books, Ebook Central, EconLit (1886 - current), ERIC (1966 - current), Ethnic NewsWatch, GenderWatch, Global Newsstream (1980 - current), Linguistics and Language Behavior Abstracts (LLBA) (1973 - current), Literature Online, Music Periodicals Database (1874 - current), PAIS Index (1914 - current), Performing Arts Periodicals Database (1864 - current), Philosopher's Index (1940 - current), ProQuest Dissertations & Theses Global, ProQuest Historical Newspapers: Louisville Courier Journal (1830 - 2000), ProQuest Historical Newspapers: The New York Times with Index (1851 - 2017), ProQuest Recent Newspapers: The Courier-Journal, ProQuest Recent Newspapers: The New York Times, PTSDpubs (1871 - current), Publicly Available Content Database, Sociological Abstracts (1952 - current), Worldwide Political Science Abstracts (1975 - current)

Table 4

Search Strategy in EBSCO & ProQuest

Search number	Search strategy	Concept block
1	AB("note-taking" OR "notetaking" OR "taking notes" OR "take notes")	Note-taking
2	AB((writ* OR handwrit*) AND (typing OR typed OR typewrit*))	Handwriting and typing
3	1 OR 2	
4	AB(recall OR memory OR retention)	Recall
5	3 AND 4	
6	Repeat 5 for title and keywords as well as abstract.	

Table 5

Studies used for search validation

Author	Year
Mueller & Oppenheimer	2014
Kodaira	2017
Bui et al.	2013
Wei et al.	2014
Mitchell & Zheng	2019
Jamet et al.	2020
Urry et al.	2021
Morehead et al.	2019
Luo et al.	2018
Slavina	2018

Table 6

Influence analysis

Authors	dfbs_intrcpt	rstudent	dffits	cook.d	cov.r	tau2.del	QE.del	hat	weight
Mueller & Oppenheimer, 2014	0.008	-0.018	0.008	0.000	1.125	0.031	52.901	0.048	4.812
Kodaira, 2017	0.202	1.021	0.202	0.041	1.040	0.026	51.224	0.038	3.766
Bui, Myers, & Hale, 2013	-0.327	-1.990	-0.321	0.096	0.926	0.019	47.937	0.025	2.478
Chiaraluce, 2019	-0.092	-0.600	-0.092	0.009	1.048	0.027	52.410	0.024	2.446
Desselle & Shane, 2018	0.220	1.494	0.219	0.047	0.989	0.024	50.083	0.023	2.260
Bargione, 1987	0.068	0.405	0.069	0.005	1.057	0.028	52.689	0.024	2.355
Schoen, 2012	-0.462	-3.802	-0.429	0.170	0.808	0.013	37.199	0.015	1.475
Beck, 2014	-0.002	-0.045	-0.002	0.000	1.036	0.027	52.899	0.014	1.406
Wei et al., 2014	0.162	1.251	0.162	0.026	1.007	0.025	51.036	0.017	1.696
Blankenship, 2016	0.127	0.604	0.127	0.017	1.080	0.028	52.291	0.038	3.848
Cubillo, 2017	-0.543	-1.994	-0.575	0.228	0.826	0.012	46.584	0.055	5.470
Crumb et al., 2020	0.197	0.998	0.197	0.039	1.043	0.026	51.303	0.037	3.739
Mitchell & Zheng, 2019	0.126	0.429	0.122	0.017	1.154	0.031	52.179	0.068	6.832
Flanagan & Titsworth, 2020	-0.024	-0.190	-0.024	0.001	1.060	0.028	52.851	0.024	2.371
Jamet et al., 2020	0.036	0.093	0.036	0.001	1.149	0.032	52.882	0.058	5.833

Artz et al., 2020	-0.102	-0.439	-0.099	0.011	1.144	0.031	52.158	0.066	6.587
Gür, 2021	0.435	2.624	0.425	0.150	0.817	0.013	44.406	0.032	3.204
Murphy, 2016	0.139	0.791	0.139	0.020	1.049	0.027	52.024	0.028	2.840
Kuipers, 2019	0.222	0.996	0.222	0.049	1.053	0.026	50.992	0.047	4.709
Urry et al., 2021	-0.169	-0.848	-0.169	0.029	1.054	0.027	51.667	0.039	3.922
Sheppard, 2015	-0.105	-0.889	-0.105	0.011	1.017	0.026	51.977	0.014	1.389
Morehead, Dunlosky, & Rawson, 2019	-0.095	-0.593	-0.095	0.009	1.053	0.027	52.403	0.027	2.688
Wood et al., 2012	-0.049	-0.388	-0.050	0.003	1.042	0.027	52.712	0.018	1.822
Duran & Frederick, 2013	0.057	0.398	0.057	0.003	1.042	0.027	52.713	0.017	1.726
Kirkland, 2016	-0.114	-0.715	-0.114	0.013	1.044	0.027	52.193	0.026	2.570
Eason, 2017	0.020	0.099	0.020	0.000	1.054	0.028	52.891	0.021	2.081
Kutta, 2017	-0.077	-0.499	-0.077	0.006	1.055	0.028	52.553	0.026	2.561
Kennedy, 2019	-0.015	-0.128	-0.015	0.000	1.069	0.028	52.877	0.027	2.679
Lalchandani, 2016	-0.023	-0.235	-0.023	0.001	1.029	0.027	52.837	0.012	1.181
Luo et al., 2018	-0.024	-0.192	-0.024	0.001	1.057	0.028	52.851	0.022	2.245
Slavina, 2018	0.006	-0.001	0.006	0.000	1.068	0.028	52.902	0.026	2.587
Tischner, 2017	-0.003	-0.058	-0.003	0.000	1.046	0.028	52.897	0.018	1.769
Devers, 2015	0.149	0.888	0.149	0.022	1.039	0.027	51.828	0.027	2.656

Notes. Influence diagnostics for an aggregated single level model.

dfbs_intrcpt = The DFBETAS statistic; effect of deleting i^{th} observation for the overall standardized mean difference.

rstudent = The (studentized) externally standardized residuals; residuals for model fit for all values except the i^{th} observation.

dffits = The DFFITS statistic; scaled measure of the change in the predicted value for the i^{th} observation from the deletion of it.

cook.d = The Cook's distance value; effect of deleting i^{th} observation on cook's D.

cov.r = The covariance ratio; measure how precision is affected by the deletion of the i^{th} observation.

tau2.del = The τ^2 deletion statistic; effect of deleting the i^{th} observation on the estimate of heterogeneity in the model.

QE.del = The Q deletion statistic; effect of deleting the i^{th} observation on the test statistic for heterogeneity.

hat = The hat value; effect of deleting the i^{th} value on the fitted values.

weight = The weight of that study in the model.

Table 7

Intercept only model

Effect	Estimate	*SE*	*t*	*df*	*p*
Intercept	0.144	0.059	2.438	28.862	0.021

Table 8

Univariate moderator tests

Model	Estimate	*SE*	*t*	*df*	*p*	*ΔR²*	Reports	Effect sizes
Delay								
Intercept	0.230	0.079	2.893	10.750	0.015	---	33	88
Same day?	-0.121	0.071	-1.697	9.690	0.122	4.874%	---	---
Review								
Intercept	0.246	0.081	3.024	10.540	0.012	---	24	61
Review?	-0.208	0.116	-1.803	15.390	0.091	2.775%	---	---
Question								
Intercept	0.075	0.113	0.659	8.940	0.526	---	21	63
Question type?	0.023	0.108	0.218	5.070	0.836	0.000%	---	---

Note. This table reports the results of a multivariate model. Same day = 1 if the outcome measure was given on the same day, 0 if it was given on a different day. Review = 1 if they did not have an opportunity to review, 0 if participants had an opportunity to review their notes. Question type = 0 for conceptual questions and 1 for factual questions.

Table 9

Multivariate Moderator Test

Effect	Estimate	*SE*	*t*	*df*	*p*
Intercept	0.468	0.098	4.750	3.068	0.017
Same day?	-0.099	0.071	-1.390	3.578	0.245
Review?	-0.414	0.109	-3.803	5.900	0.009
Question type	-0.006	0.132	-0.044	3.412	0.968

Note. This table reports the results of a multivariate multilevel model. Same day = 1 if the outcome measure was given on the same day, 0 if it was given on a different day. Review = 1 if participants did not have an opportunity to review, 0 if participants had an opportunity to review their notes. Question type = 0 for conceptual questions and 1 for factual questions. Reports = 14, effect sizes = 41, R^2 = 24.560%.

Table 10

Exploratory univariate model results

Effect	Estimate	*SE*	*t*	*df*	*p*	*ΔR²*	Reports	Effect sizes
Model 1						0.000%	32	84
Intercept	0.104	0.058	1.788	21.684	0.088	---	---	---
Live lecture?	0.076	0.145	0.523	3.519	0.632	---	---	---
Model 2						0.000%	31	86
Intercept	0.229	0.068	3.384	6.558	0.013	---	---	---
Random?	-0.124	0.105	-1.185	11.201	0.260	---	---	---
Model 3						0.000%	33	88
Intercept	0.064	0.083	0.764	12.130	0.460	---	---	---
Publication Type: Journal	0.112	0.122	0.915	24.680	0.369	---	---	---
Publication Type: Conference Pres.	0.334	0.085	3.918	1.292	0.114	---	---	---
Model 4						0.000%	33	88
Intercept	0.123	0.066	1.859	25.088	0.075	---	---	---
Environment: Secondary	0.225	0.073	3.084	2.404	0.072	---	---	---
Environment: Other	0.047	0.075	0.623	1.611	0.610	---	---	---
Model 5						0.000%	32	86
Intercept	0.151	0.090	1.683	16.667	0.111	---	---	---
Question Type: Short Answer	-0.086	0.123	-0.697	15.918	0.496	---	---	---

Effect	Estimate	*SE*	*t*	*df*	*p*	*ΔR²*	Reports	Effect sizes
Question Type: Other	0.055	0.138	0.398	11.196	0.698	---	---	---
Model 6						0.000%	32	84
Intercept	0.181	0.134	1.349	2.772	0.277	---	---	---
Recording: Other	-0.013	0.134	-0.094	2.776	0.932	---	---	---
Recording: Yes	-0.081	0.147	-0.553	3.588	0.613	---	---	---
Model 7						*42.56%	32	84
Intercept	0.155	0.039	3.995	18.988	0.001	---	---	---
Audio	-0.605	0.302	-2.004	1.135	0.271	---	---	---
Model 8						3.027%	31	83
Intercept	0.111	0.087	1.271	12.324	0.227	---	---	---
Delay	0.062	0.089	0.693	10.797	0.503	---	---	---
Model 9						0.000%	19	53
Intercept	0.212	0.055	3.837	13.719	0.002	---	---	---
Delay minutes	0.000	0.000	-0.009	1.586	0.994	---	---	---
Model 10						0.000%	32	84
Intercept	0.124	0.091	1.363	13.711	0.195	---	---	---
Distractor	-0.022	0.100	-0.223	19.433	0.826	---	---	---
Model 11						0.000%	7	24
Intercept	-0.011	0.080	-0.134	1.002	0.915	---	---	---
Intention	0.252	0.135	1.868	1.887	0.210	---	---	---

Effect	Estimate	*SE*	*t*	*df*	*p*	ΔR^2	Reports	Effect sizes
Model 12						0.000%	33	88
Intercept	0.096	0.138	0.692	4.401	0.524	---	---	---
Transcription Capacity	0.059	0.141	0.414	5.354	0.695	---	---	---
Model 13						0.000%	14	51
Intercept	0.132	0.066	2.005	6.090	0.091	---	---	---
Overall Attrition	0.041	0.150	0.275	2.644	0.803	---	---	---
Model 14						4.339%	33	88
intercept	0.198	0.090	2.203	17.037	0.042	---	---	---
Data type: Dichotomized	-0.146	0.094	-1.545	22.200	0.136	---	---	---

Notes.This table reports the results of multiple univariate models. Same day = 1if the outcome measure was given on the same day, 0 if it was given on a different day. Live lecture =1 if the lecture administered was from a live lecturer. Random = 1 if the participants were randomly assigned to treatment and control groups. Publication Type = 0 for thesis and dissertations, 1 for journals, and 2 for conference presentations. Environment = 0 for university, 1 for secondary, and 2 for other. Question type = 0 for multiple choice, 1 for short answer and 2 for other. Recording = 0 for no, 1 for other and 2 for yes. Audio = 0 audio + visual and 1 for audio only. Delay = 0 for no delay and 1 for a delay. Delay minutes = 1 for a 1 minute increase in the delay. Distractor = 0 for no distractor task and 1 for yes. Intention = 0 for no intention and 1 for yes. Transcription capacity = 0 for true and 1 for false. Attrition = 1 for each individual who was dropped from the study. Data type = 0 for continuous and 1 for dichotomized.

* = 76 out of 84 effect sizes (90.5%) were audio + visual input and 8 were audio only condition.

APPENDIX

A. CODE

https://github.com/timothyslau/dissertation

CURRCULUM VITA

Timothy S. Lau
(410) 417-8760 t@lau.dev

EDUCATION

University of Louisville, Louisville, KY

PhD Educational Psychology, Measurement, & Evaluation **2022**

Dissertation: "The effect typewriting vs. handwriting lecture notes on learning: A systematic review and meta-analysis"
Advisor: Jeffrey C. Valentine, Ph.D.

Johns Hopkins University, Baltimore, MD

M.S. Technology for Educators **2011**

Additional Certificate: Data Driven Decision Making
Advisor: John Castellani, Ph.D.

University of Nevada, Las Vegas

B.A. Honors in Psychology **2009**

Area of Concentration: Cognition, Mathematical
Honors Thesis: "Estimation Verification"
Advisor: Mark H. Ashcraft, Ph.D.

AWARDS

University of Louisville, Louisville, KY

- Tuition Waiver **2015-2016**
- APA, Advanced Training Institute, Longitudinal Structural Equation Modeling **2015**
- APA, Advanced Training Institute, Non-Linear Modeling **2014**
- Samuel's Scholarship for Academic Achievement **2012-2013**

Johns Hopkins University, Baltimore, MD

- BGE Scholarship **2010-2011**

University of Nevada, Las Vegas

- Magna Cum Laude **2009**
- Department Honors **2008 – 2009**
- Deans Honor List **Spring, Fall 2008 – Fall 2009**

Coronado High School, Henderson, NV

- Millennium Scholarship **2003 – 2009**
- Advanced Graduate, Coronado High School **2003**
- Scholar Athlete, Coronado High School **2002 – 2003**

PUBLICATIONS

- Lau, T. S. (2022) The effect of typewriting vs. handwriting lecture notes on learning: A systematic review and meta-analysis. 2022
- Ghosal, S., Lau, T. S., Gaskins, J., Kong, M. (2020). A hierarchical mixed effect hurdle model for spatiotemporal count data and its application to identifying factors impacting health professional shortages. *Journal of the Royal Statistical Society.* 2020
- Devlin, L. A., Lau, T. S., Radmacher, P. G. (2017). Decreasing total medical exposure and length of stay while completing withdrawal for neonatal abstinence syndrome during neonatal hospital stay. 2017
- Pittard, C. M., Pössel, P., Lau, T. S. (2017). Inferential style, school teachers, and depressive symptoms in college students. *International Journal of Emotional Intelligence.* 2017
- Valentine, J. C., Tanner-Smith, E. E., Pustejovsky, J. E., Lau, T. S. (2016). Between-case standardized mean difference effect sizes for single-case designs: a primer. *The Campbell Collaboration.* 2016
- Valentine, J. C., Wilson, S. J., Rindskopf, D., Lau, T. S., Tanner-Smith, E. E., Yeide, M., LaSota, R., & Foster, L. (2016). Synthesizing evidence in public policy contexts: The challenge of synthesis when there are only a few studies. *Evaluation Review.* 2016
- Valentine, J. C., Aloe, A. M., & Lau, T. S. (2015). Life after NHST: How to describe your data without "p-ing" everywhere. *Basic and Applied Social Psychology.* 2015

- Valentine, J. C., Pigott, T. D., & Lau, T. (2013). Systematic reviewing and meta-analysis In J. Wright (Ed.) International Encyclopedia of the Social and Behavioral Sciences, 2nd ed. — 2013
- "Treatment Foster Care for Reducing Internalizing and Externalizing Symptoms in Foster Youth: A Systematic Review and Meta-Analysis Manuscript, The Cochrane Collaboration, Louisville, KY — **2013**
- Intervention Reports What Works Clearinghouse, Department of Education, United States Federal Government. — **2013 – 2015**
- Lau, T. S. (2009) Estimation Verification. — **2009**

PRESENTATIONS

- Lau, T. S. & Upton, T. (2015). Visualizing non-parametric alternatives for statistics education. Southeast Evaluation Association. Tallahassee, FL. — **2015**
- Valentine, J. C. & Lau, T. S. (2015). Synthesizing evidence: Synthesis methods for evidence clearinghouses. Society for Research on Educational Effectiveness, Washington DC. — **2015**
- Valentine, J. C., Konstantopoulos, S., Lau, T. S. (2014). Protocol for systematic review: comparing alternatively certified to regularly certified teachers: A systematic review and meta-analysis. — **2014**
- "A Visual Guide to Proper Nonparametric Method Selection" University of Louisville Graduate Student Research Symposium, Louisville KY. — **2014**
- "The Differences Between Writing and Typing While Note-taking" Graduate Research Symposium, Louisville, KY — **2013**
- "Notation Modalities" Graduate Research Symposium, Louisville, KY — **2012**
- "Omnibus University Outcomes & Implications" Data analysis, Utah Valley University, Orem, UT — **2011**
- "Online Survey Development" Consulting for MSET Conference, Baltimore, MD — **2011**
- "Improving Education" Johns Hopkins University, School of Education, Scholars Award Ceremony, Baltimore, MD — **2009**
- "Estimation Verification" Poster presented to the Honors College for Department Honors, Las Vegas, NV — **2009**
- "Estimation Verification: Number-Space Representation" Thesis written for the Honors College for Department Honors, Las Vegas, NV — **2009**
- "Variations in Estimation" WPA Convention, Portland, OR — **2008**
- Estimation Verification Psi-Chi Poster Conference, Las Vegas, NV — **2008**

TEACHING EXPERIENCE

University of Louisville, Louisville, KY

Graduate Mentor — **2013 – 2016**

Mentored and instructed undergraduate students in statistics, statistical programming, research methodology, and preparation for graduate school.

Graduate Teaching Assistant — **2012 – 2013**

Assisting instruction in courses on structural equation and higher linear modeling.

Guest Lecturer — **2012**

Intelligence and child development.

Center for Talented Youth, Johns Hopkins University, Baltimore, MD

Computer Science Instructor — **2011 – 2012**

Teaching, recording lectures for, and designing online courses in Scratch programming and Web Design.

Johns Hopkins University, Baltimore, MD

Private Tutor — **2010 – 2011**

Web 2.0 tools, media design, e-learning environments, web-based communication tools, data gathering, data manipulation, data display, and HTML coding.

Clark County School District, Las Vegas, NV

Substitute Teacher — **2008 – 2010**

K-12 "All Classes". Taught and interacted with students, developed lesson plans, and graded written work.

University of Nevada, Las Vegas

Private Tutor **2007 – 2010**

Statistical Methods in Psychology, College Algebra, and General Chemistry. Developed lesson plans and overall learning structure.

Research Assistant Trainer **2007 – 2009**

Mathematical Cognition Lab, taught and trained new research assistants.

Private Homes, Henderson, NV

Private Instructor **2007**

ABC Behavioral Program for Autism

Cherry Valley Elementary, Polson, MT

Tutor **2005 – 2006**

"English Literacy Program" (Salish/Kootenai Indian Reservation). Developed reading ability and phonetic reading comprehension.

RESEARCH EXPERIENCE

Educational Psychology, Measurement and Evaluation Research Lab, University of Louisville, Louisville, KY

Research Lab Manager **2012 – 2016**

Created and managed a research lab, of up to 8 people, consisting of graduate and undergraduate students researching robust methods of statistical inference, regression assumption simulation, advanced statistical programming, Bayesian regression and meta-analysis, non-linear modeling, and student note-taking.

What Works Clearing House, Department of Education, Federal Government, Washington, DC

Research Assistant **2012 – 2016**

Research synthesis, meta-analysis, data analysis, contributor to Dept. of Ed. What Works Clearinghouse, contributor to The Cochrane Collaboration article publication, article review, database management, R programing, and article retrieval.

Campbell Collaboration, Oslo, Norway

Research Assistant **2012 – 2013**

Systematic Review and meta-analysis for a grant funded study of alternative route to licensure for teachers.

Dr. Jill L. Adelson, University of Louisville, Louisville, KY

Research Assistant **2012 – 2013**

Research synthesis, literature review, manuscript review, data analysis, report writing, article reviews, and article retrieval.

Institutional Research and Information, Utah Valley University, Orem, UT

Research Analyst **2011 – 2012**

Research, data analysis, and organizational upkeep on data gathering for university administration.

Center Research and Reform in Education, Johns Hopkins University, Baltimore, MD

Research Intern **2011**

Worked collaboratively to assist the research and evaluation coordinator as needed. Work included editing, designing, and developing a research database, reviewing journal articles and meta-analysis, and writing reports for research criteria met for inclusion in a publication of a synthesis of meta-analysis in Computer Assisted Instruction.

Dr. Deborah Carran, Johns Hopkins University, Baltimore, MD

Research Intern **2010 – 2011**

Worked on organizing data, data analysis, and traveling to schools to gather qualitative and quantitative data. Assisted in survey, evaluation, and experiment design, development, and implementation.

Dr. John Castellani, Johns Hopkins University, Baltimore, MD

Research Intern **2010 – 2011**

Worked on managing and analyzing large data sets for publication.

Living Classrooms Foundation, Baltimore, MD

Research Consultant **2010 – 2011**

Consulted on and contributed to the building of online databases and web platform development.

Dr. Mariale Hardiman, Johns Hopkins University, Baltimore, MD

Research Intern 2010

Contributed research on the development of a metric for creativity. Contributed current research for updating a chapter of a book. Developed an online research database that could be shared and collaboratively receive input.

Center for Technology in Education, Johns Hopkins University, Baltimore, MD

Research Intern 2010

Worked collaboratively to assist the research and evaluation coordinator as needed. Work included planning and constructing evaluations, surveys and then analyzing and interpreting the data. Wrote a literature review for a multi-year longitudinal grant proposal.

Math Cognition Lab, University of Nevada, Las Vegas

Research Assistant 2007 – 2009

Performed multiple experiments involving eye-tracking equipment. Wrote supporting documents and trained others on how to operate the equipment as needed.

Memory Lab, University of Nevada, Las Vegas

Researcher Assistant 2008

Simultaneously ran multiple subjects and experiments.

FRIAS Project, University of Nevada, Las Vegas

Psychology Lab Assistant Manager 2007

Assisted in setting up and performing experiments for the FRIAS Project—a yearlong grant funded study, examining number sense and mathematical proficiency of children 1st through 5th grade.

MEMBERSHIPS

- American Statistical Association **2015 – 2019**
- South Eastern Evaluation Association **2013 – 2016**
- American Evaluation Association **2013 – 2016**
- GSA, Graduate Student Association, Vice President **2013 – 2016**
- Kentucky Academy of Science **2012 – 2016**
- GSC, Graduate Student Council, Program Representative **2012 – 2016**
- PRIMR, Public Responsibility in Medicine and Research **2012 – 2016**
- The Cochrane Collaboration, Researcher **2012 – 2016**
- Public Responsibility in Medicine and Research **2012 – 2016**
- IRB review board, University of Louisville **2012 – 2016**
- Budget Advisory Committee, University of Louisville **2012 – 2016**
- Financial Health Committee, University of Louisville **2012 – 2016**
- Technology/Online Subcommittee of the 21st Century UofL Initiative **2012 – 2016**
- Pi Lambda Theta, Intl. Honor Society in Education **2010 – 2016**
- FEA, Future Educators Association **2010 – 2016**
- Learning and the Brain Society **2010 – 2016**
- IMBES, Intl. Mind Brain and Education Society **2010 – 2016**
- AERA, American Educational Research Association **2010 – 2016**
- Phi Delta Kappa Intl., Professional Education Association **2010 – 2016**
- NHMSS, National Honor & Merit Scholars Society **2009 – 2016**
- Magna Cum Laude, The National Honor Society **2009 – 2016**
- American Psychological Science **2008 – 2012**
- Western Psychological Association **2008 – 2012**
- SIFE, Students In Free Enterprise, Vice President **2009 – 2010**
- Golden Key, Intl. Honor Society, Fundraising Chair **2009 – 2010**
- Psi-Chi, National Honor Society, Finance Chair **2007 – 2010**
- UNLV Psychology Club, Finance Chair **2007 – 2010**
- OUMP, Outreach Undergraduate Mentoring Program **2007 – 2010**

COMMUNITY SERVICE

- Graduate Student Orientation Panel, UofL **2012 – 2016**
- Louisville Urban League **2012**
- Living Classrooms Foundation **2010 – 2012**
- LBS, Learning and the Brain Society **2010 – 2012**
- FEA, Future Educators Association **2010 – 2012**
- SIFE service projects, teaching financial literacy **2009 – 2011**
- OUMP, Outreach Undergraduate Mentoring Program **2007 – 2011**
- Candle Lighters Volunteer, Childhood Cancer Foundation **2007 – 2010**
- UNLV Disability Resource Center **2007 – 2010**

SOFTWARE PROFICIENCIES

- R
- S
- Python
- Julia
- Stata
- MATLAB
- SPSS
- HLM
- SAS
- SQL
- E-Prime
- LaTeX
- HTML/XHTML I
- CSS
- Google Apps
- Microsoft Office Suite
- EndNote Syntax
- Markdown

ACADEMIC REFERENCES

Dr. Jeffrey C. Valentine, Ph.D.
Professor of Educational and Counseling Psychology
University of Louisville
jeff.valentine@lousiville.edu
(502) 852-3830

Dr. David Roelfs, Ph.D.
Associate Professor of Sociology
University of Louisville
david.Roelfs@lousiville.edu
(502) 852-8038

Dr. Stephan Taeger, Ph.D.
Professor of Religion
Brigham Young University
stephan_taeger@byu.edu
(801) 422-2735

Dr. Maiying Kong, Ph.D.
Professor of Bioinformatics and Biostatistics
University of Louisville
maiying.kong@lousiville.edu
(502) 852-3830

Rebecca Higgins, CIP
Director of Institutional Review Board
University of Louisville
rebecca.higgins@lousiville.edu
(502) 852-5188

Dr. Alan Cheung, Ph.D.
Associate Professor at the Center for Research and Reform in Education
Johns Hopkins University
acheung@jhu.edu
(410) 616-2410

Dr. Russell T. Hurlburt, Ph.D.
Professor of Psychology
University of Nevada, Las Vegas
russ@unlv.nevada.edu
(702) 895-0194

Dr. Mark H. Ashcraft, Ph.D.
Department Chair of Psychology
Advisor – Department Honors Thesis
University of Nevada, Las Vegas
mark.ashcraft@unlv.edu
(702) 859-0175