Contemporary Educational Psychology 68 (2022) 102025

Contents lists available at ScienceDirect

Contemporary Educational Psychology

journal homepage: www.elsevier.com/locate/cedpsych

The effect of notetaking method on academic performance: A systematic review and meta-analysis

Daniel Voyer *, Scott T. Ronis, Narissa Byers

Department of Psychology, University of New Brunswick, Canada

A R T I C L E I N F O

Keywords:
Notetaking
Meta-analysis
School achievement
Cognitive performance

A B S T R A C T

The present meta-analysis aimed to synthesize the extant research on the influence of longhand (written) versus digital notetaking methods, unconfounded by distractions, on performance, and to identify key potential moderators of such effects. After a systematic literature search, we obtained 77 effect sizes from 39 samples in 36 articles and conducted a multilevel meta-analysis complemented with robust variance estimation. Overall, results showed a mean effect size (mean estimated g = -0.008, 95% *CI*: −0.18, 0.16) that was not significantly different from zero, suggesting no effect of notetaking approach. Moderator analyses, justified by effect sizes with significant heterogeneity, demonstrated only three significant moderators referring to the topic covered, the learning objectives, and the duration of the material. Overall, however, the present results suggest 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. Nevertheless, more research is required to identify moderators that might account for variability in the findings. Work using large representative samples of all ages is particularly needed.

From primary school through post-secondary education, students are typically expected to attend to lectures, take notes on what they see and hear, and study the material for subsequent assignments or examinations. With technology becoming ubiquitous and expansive, instructors have navigated its use in classrooms, particularly for notetaking of lectures. Indeed, many instructors likely have discouraged broadly the use of technology because they believe it distracts from class engagement and student learning. Furthermore, many have questioned the effectiveness of digital notetaking relative to longhand notetaking (i.e., students taking notes using pen and paper) on learning and scholastic achievement even if we disregard the influence of distraction or other factors (e.g., course or lecture topic, learning objectives, duration of presented material) that may affect the link between notetaking methods and learning or achievement (Goodwin, 2018; Kennedy, 2019; Mueller & Oppenheimer, 2014). However, other educators and many students perceive that the benefits of digital notetaking outweigh potential costs (Jansen et al., 2017). Although several studies have been conducted, empirical research findings to date comparing notetaking approaches have not led to a clear understanding of recommended directions for educators or students.

1. Digital notetaking and distraction

In what appears to be the only existing meta-analysis on the effect of notetaking method on learning performance, Allen, Lefebvre, Lefebvre, and Bourhis (2020) aggregated data from 14 studies (24 estimates). Their results showed poorer performance for digital compared to longhand notetaking (average r = -0.142). However, the Allen et al. meta-analysis did not separate the influence of distraction from the notetaking method. Specifically, they acknowledged that digital notetaking also allows the influence of external distractions (e.g., email, web browsing) on learning performance. In contrast, longhand notetaking would only be affected by incidental distractors (e.g., students arriving late, other students talking, one's own attention lapses) and, even then, these factors would similarly affect digital notetaking. Therefore, the meta-analysis by Allen et al. confounds notetaking method with the presence of distractions. In addition, the search parameters that guided the meta-analysis were narrow, focusing exclusively on the classroom experience and on college students. Accordingly, there has been no systematic meta-analysis comparing performance under longhand or digital notetaking that is not contaminated by the influence of distractions and including a broad range of testing situations (beyond the

* Corresponding author at: Department of Psychology, University of New Brunswick, Box 4400, Fredericton, New Brunswick E3B 5A3, Canada.
E-mail address: voyer@unb.ca (D. Voyer).

https://doi.org/10.1016/j.cedpsych.2021.102025

Available online 8 November 2021
0361-476X/

college classroom) and age groups. Our purpose was, therefore, to synthesize the extant research on the influence of notetaking method in all its forms to assess possible positive or negative consequences and to identify whether there are key potential moderators of these effects. Our hope was that such a meta-analysis might clarify the outcome of relevant studies and promote clearer guidance for instructors.

2. Notetaking functions

In general, there is a robust benefit of notetaking on performance, as it typically serves process-related and product functions (Luo et al., 2018). Among the process-related functions, researchers have identified the promotion of attention and encoding of presented material as well as the roles of complex cognitive demands in memory and performance. In particular, after attending to presented material, notetaking involves a range of generative (e.g., summarizing, paraphrasing) and nongenerative (i.e., verbatim transcription of information) approaches (Kodaira, 2017; Stefanou et al., 2008). Consistent with the encoding hypothesis (Di Vesta & Gray, 1972; Kiewra et al., 1991; Tulving & Thomson, 1973), researchers have considered that generative approaches involve more cognitive processing than a nongenerative approach and, as a result, provide benefits of deeper encoding (for efficient recall) and learning material.

Beyond considering the process of notetaking on learning and academic performance, a product is created, usually in a visual, written format (lecture notes), that can generally aid in rehearsal, organization, and studying material. Indeed, students often review and study their notes to be able to learn target information (McCann et al., 2019). According to the external-storage hypothesis (Di Vesta & Gray, 1972; Kiewra et al., 1991), by doing so, what was originally attended to and recorded in class is processed again and perhaps in a more complex manner. Overall, studying recorded notes does lead to improved short-term (immediate) recall (through working memory) and increased academic performance relative to not studying (Di Vesta & Gray, 1972). Furthermore, with more complex processing and application of the material, information may be stored in long-term memory (Lockhart & Craik, 1990).

3. Notetaking approaches

Notetaking typically is completed through either a longhand format or a digital format with computers, tablets, or other technological devices. In particular, digital notetaking on laptops or tablets has been used increasingly, which has engendered questions about the effects of this form of notetaking on attention, memory, learning, recall, and performance. Because individuals generally type notes faster than when writing (Aragón-Mendizábal et al., 2016), typed notes tend to be nongenerative and generally more comprehensive. Furthermore, given the speed benefit and association in general between the amount of notes taken and amount of encoded information (Bui et al., 2013; Fiorella & Mayer, 2017), some researchers (e.g., Bui et al., 2013; Schoen, 2012) have suggested that individuals do better on performance tasks with digital notetaking. However, there have been inconsistent findings in the literature, perhaps due to variations in class topics or assessment methods.

In comparing the two approaches, it is important to consider what is involved in each. When presented with audiovisual information through a lecture, students must divide their attention between the presented information and writing or typing notes. Researchers have generally found more task switching and divided attention in digital notetaking than in written notetaking, as students are more prone to distractions (e.g., web surfing, texting) (Mueller & Oppenheimer, 2014; Waite et al., 2018). Alternatively, because written notetaking has been shown to produce greater cognitive load, some researchers have suggested that with multiple processing involved in notetaking (e.g., listening, summarizing, paraphrasing), students who take longhand notes perform better on overall recall tasks (e.g., Aragón-Mendizábal et al., 2016; Mueller & Oppenheimer, 2014; Olive & Barbier, 2017; Patterson & Patterson, 2017). However, many studies have also found either the opposite (e.g., Fiorella & Mayer, 2017; Jackson, 2016) or that notetaking was not associated with performance (e.g., Kutta, 2017). An explanation for some studies finding different levels of performance may center on unique learning strategies that students use in one approach or another (Jansen et al., 2017). For example, students using longhand notetaking are more likely to use spatial strategies (e.g., creating maps, drawing), whereas students using digital notetaking more often use verbal strategies (e.g., making lists or outlines) (Fiorella & Mayer, 2017). In breaking down recall tasks between short- and long-term tasks, Beck (2014) found that students who take longhand notes scored higher on both types of tasks than students who take typed notes.

4. Role of depth of processing

One obvious aspect that distinguishes written and digital notetaking pertains to mechanistic processes. For example, the two approaches involve different physical movements, use of physical objects, production of concrete products, freedom to or ease of ability to draw, color, scale, or work in non-linear fashion. Of course, these mechanistic constraints vary also with specific digital platforms and with individual differences. Interestingly, although some researchers have examined mechanistic processing in notetaking (Kodaira, 2017; Piolat, Olive, & Kellogg, 2005), most research has focused on cognitive processes (e.g., depth of processing, executive attention, working memory). Consistent with this literature, we suggest that hypothetical differences between longhand notetaking and digital notetaking are due to variations in cognitive processes. In particular, some researchers have found generally a positive association between the generative processing associated with handwriting and performance (e.g., Davis & Hult, 1997) as well as a link between deeper processing in working memory with handwritten notes and performance (Bui & Myerson, 2014; Kiewra, 1985; Mueller & Oppenheimer, 2018; Peverly et al., 2007). In this context, our purpose to consider studies that equalize distractions across notetaking conditions has more to do with encoding (attentional) processes than with freedom or constraints of physical movements.

In general, substantial variability in terms of empirical methods and results limit our ability to draw definitive conclusions. With this variability in mind and in keeping with best recommended practices in meta-analysis (Johnson & Hennessey, 2019), we included a limited number of pre-specified potential moderators. In addition, still following recommended best practices, the protocol was pre-registered and can be consulted at osf.io/zm8nh (Registration https://doi.org/10.17605/OSF.IO/V6GJZ).

5. Potential moderators

5.1. Sample factors

In examining longhand and digital notetaking approaches, it is important to consider that some sample factors potentially moderate effects on academic performance. In particular, demographic information (e.g., socioeconomic status, ethnicity) and other sample characteristics (e.g., clinical versus general community sample) are considered to have important direct or indirect effects on academic performance (Gurley, 2018; Richards et al., 2016). For sample characteristics, although studies comparing different approaches generally use university samples because students often select relatively evenly among the notetaking approaches (Morehead et al., 2019), some studies (e.g., Igo et al., 2006; Jackson, 2016; Jiang et al., 2018) have included other age groups and level of schooling in examining performance. Therefore, we included age, which is a proxy for level of schooling, as a moderator.

5.2. Measurement factors

Design-related factors. Because experimental studies have varied based on the methods and tasks that were used, there are potential factors that differentially affect learning and academic performance with longhand and digital notetaking (Luo et al., 2018). At the forefront, compared to correlational or quasi-experimental studies, relatively few studies have employed within-subject or controlled experimental designs (e.g., Artz et al., 2020) to assess method-related explanations (e.g., longhand versus digital notetaking) while accounting for participant characteristics (e.g., students selecting one notetaking method over another). In addition, some evidence suggests that congruency between notetaking approach and assessment approach affects performance (Barrett et al., 2014; Olive & Barbier, 2017). Overall, it is essential to consider the degree of control in research designs when drawing conclusions. Accordingly, we coded whether each study was from an experiment where the notetaking method was manipulated directly, or a quasi-experiment in which it was incidental (i.e., a self-selected method for participants).

The role of distraction. It is likely that controlling the presence of distractions during the notetaking session would affect performance. Indeed, several experimental and quasi-experimental studies have identified distractions as negatively affecting the amount of encoded information during notetaking (Allen et al., 2020). Accordingly, we noted whether or not the study involved distraction and whether the distraction was manipulated or incidental (e.g., distractions incidental to a regular classroom, such as where students might arrive late, cough, or talk to each other). Inclusion of this moderator addressed directly the confound of notetaking method and distraction in the Allen et al. (2020) meta-analysis, reflecting a central purpose of the present meta-analysis.

Factors Relevant to Assessment. We would be remiss to present potential moderators of the effect of notetaking method without considering the course or lecture topic. Specifically, the topic of the courses likely guides the type of assessment that is applied, with some areas (e.g., history) lending themselves to a heavy factual memory load, whereas text analysis in literature likely calls on a heavy conceptual load. In addition, different course topics presumably carry distinct learning objectives. For example, learning objectives in literature tend to be about themes rather than about rote memory. Accordingly, we coded topics in as fine a grain as possible and categorized the learning objectives to capture subtle differences.

Although originally disregarded in the pre-registration of this work, we discovered as we retrieved the research that the literature afforded an investigation of variables relevant to the assessment methods. Specifically, we considered test type (recall, recognition, or both) as a potential moderator because recall tasks might be presumed to show a larger effect of notetaking methods relative to recognition tasks given greater inherent difficulty (Cabeza et al., 1997). Although results reported by Blankenship (2016) fit with this expectation, as this author found a greater advantage for handwritten over typed notes for recall than for recognition, these results were not replicated by Aragón-Mendizábal et al. (2016) or Smoker et al. (2009). We also included material type (factual/memory, conceptual, or mix) because it has received much attention in replication attempts since its inclusion in the Mueller and Oppenheimer (2014) study. This variable essentially reflects a crucial dimension in assessment because it refers to whether comprehension or applications (conceptual) were considered or if it contained primarily rote memory (factual) questions. In their study, Mueller and Oppenheimer found that handwritten notes produced better performance for conceptual than for factual material. However, replication of this finding has been variable (see Table 1). The current meta-analysis therefore examined more closely the potential role of test type and material type in the broad literature.

6. Current meta-analysis

The purpose of the present meta-analysis was to fill a gap in the literature by determining whether the method used in notetaking during a course lecture or relevant experimental tasks affects performance. In addition, we documented potential factors that moderate this effect to guide future basic research as well as classroom applications, following approaches that promoted transparency and reproducibility. To our knowledge, aside from the limited analysis conducted by Allen et al. (2020), there is no comprehensive meta-analysis examining whether there is an advantage to taking notes by longhand or by digital means. This is a puzzling omission considering the anecdotal support offered by instructors to either ban computers and other means of digital notetaking in their classroom or to encourage that form of notetaking. It is thus quite clear that our meta-analysis fills a gap with crucial practical implications.

Our main purpose was to isolate the potential role of distraction in digital notetaking by considering only studies where this factor was equated across conditions, either through experimental manipulations or instructions. Therefore, we aimed to find a measure of the role of notetaking methods where handwritten and digital approaches are equalized on distraction. Although some might argue that we cannot truly remove distractions from digital methods because they are "inherent" to them, our literature search provided examples of studies where this was achieved. It is, therefore, possible to fully control distractions in a laboratory setting (e.g., Fiorella & Mayer, 2017; Lin & Bigenho, 2011), whereas the natural classroom environment only allows us to equate the notetaking methods to overall level of distraction inherent to the classroom environment. From this perspective, the present meta-analysis examined whether handwritten and digital methods produce differential performance when the level of distraction is equated across notetaking methods so that digital devices are used as a pure notetaking method, and nothing else (i.e., not devices used to check one's email or social media pages).

In many cases, multiple outcome variables or experimental conditions relevant to our selected moderators came from the same sample of participants. In such cases, the non-independence of effect sizes would invalidate results obtained with a classic random-effect meta-analysis. In contrast, multilevel meta-analysis is designed to deal with the non-independence of measurement and hierarchical nature of the data (Raudenbush & Bryk, 2002; Van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2013, 2015). Therefore, by using a multilevel meta-analytic approach, we preserved all relevant effect sizes without the need for arbitrary selection or combination of estimates. In addition, to ensure that correlated variance estimates were considered, we also applied robust variance estimates. It is worth noting that Allen et al. (2020) did not address non-independence among their effect sizes even though it existed (24 effect sizes from 14 studies). This raises the possibility that their overall estimate was inaccurate and that their underestimated standard errors increased the risk of Type 1 errors, especially considering the small number of samples involved (Tipton & Pustejovsky, 2015). Accordingly, with our approach to the meta-analysis, we made our search more comprehensive than Allen et al. (2020) and addressed statistical limitations in their analysis.

We placed no restrictions on the range of years covered by our search, the language of publication, or on the age of participants to obtain a comprehensive sample of available research. In addition, we obtained some unpublished work and contacted researchers to obtain relevant information when it was not presented in the retrieved literature. Our analysis should therefore provide a valid reflection of the available data.

Table 1
Studies entered in the meta-analysis.

Authors	Year	Age	NH	ND	Topic	Interval	Material	Design	Study Type	Distr	Duration	Sample	d
Aragón-Mendizábal et al.	2016	19.2	211	40	Mem	Min	FM	B	Q	Nat	<=10 min	Typical	−5.57*
Aragón-Mendizábal et al.	2016	19.2	211	40	Mem	Min	FM	B	Q	Nat	<=10 min	Typical	−5.43*
Aragón-Mendizábal et al.	2016	19.2	211	40	Mem	Min	FM	B	Q	Nat	<=10 min	Typical	2.58
Aragón-Mendizábal et al.	2016	19.2	211	40	Mem	Min	FM	B	Q	Nat	<=10 min	Typical	3.92
Artz et al.	2020	20.0	215	215	SS	Days	Mix	W	E	Nat	Semester	Typical	−0.12
Bennett	2018	12.5	13	15	Math	Days	Mix	B	Q	Nat	Semester	Typical	0.53
Blankenship	2016	19.5	52	34	SS	Min	Mix	B	E	Nat	<=20 min	Typical	−0.46
Blankenship	2016	19.5	52	34	SS	Min	Mix	B	E	Nat	<=20 min	Typical	−0.16
Bui et al.	2013	19.2	40	40	Lec	No	FM	B	E	Nat	<=20 min	Typical	0.32
Desselle & Shane	2018	25.0	43	43	OT	Days	Mix	B	Q	Nat	Days	Typical	−0.44
Duran & Frederick	2013	22.0	36	36	SS	No	CC	B	E	Nat	<=20 min	Typical	−0.67
Eason	2017	19.1	28	29	SS	No	Mix	B	E	Nat	<=30 min	Typical	−0.35
Eason	2017	19.1	26	26	SS	Week	Mix	B	E	Nat	<=30 min	Typical	0.62
Fiorella & Mayer	2017	19.6	32	31	OT	No	FM	B	E	no	<=10 min	Typical	0.40
Fiorella & Mayer	2017	19.6	32	31	OT	No	FM	B	E	no	<=10 min	Typical	1.13
Fiorella & Mayer	2017	19.6	32	31	OT	No	CC	B	E	no	<=10 min	Typical	1.49
Fulton et al.	2011	19.0	18	36	TC	No	NR	B	E	Man	30–60	Typical	−0.58
Hembrooke & Gay	2003	19.0	22	22	Hum	No	NR	B	Q	Nat	30–60	Typical	−0.67
Hembrooke & Gay	2003	19.0	22	22	Hum	No	NR	B	Q	Nat	30–60	Typical	−0.67
Hembrooke & Gay	2003	19.0	22	22	Hum	No	NR	B	Q	Nat	30–60	Typical	−0.56
Hembrooke & Gay	2003	19.0	22	22	Hum	No	NR	B	Q	Nat	30–60	Typical	−0.50
Horwitz	2017	19.0	12	12	SS	Min	FM	B	E	Nat	<=30 min	Typical	0.78
Igo et al.	2006	13.9	15	15	OT	No	FM	W	E	Nat	<=10 min	Atypical	−0.39
Igo et al.	2006	13.9	15	15	OT	Days	FM	W	E	Nat	<=10 min	Atypical	−0.36
Igo et al.	2006	13.9	15	15	OT	No	FM	W	E	Nat	<=10 min	Atypical	−0.09
Jackson	2015	13.5	59	59	SS	Days	NR	B	Q	Nat	Semester	Typical	0.62
Jamet et al.	2020	19.0	97	90	SS	No	FM	B	Q	Nat	<=30 min	Typical	−0.32
Jamet et al.	2020	19.0	97	90	SS	No	CC	B	Q	Nat	<=30 min	Typical	0.02
Kennedy	2019	25.0	60	71	OT	Days	NR	B	Q	Nat	Semester	Typical	−0.33
Kirkland	2016	19.0	52	53	SS	Min	Mix	B	Q	Nat	<=20 min	Typical	0.26
Kodaira	2017	19.4	40	40	Hum	Min	FM	B	E	Nat	<=20 min	Typical	−0.50
Kodaira	2017	19.4	40	40	Hum	Min	CC	B	E	Nat	<=20 min	Typical	−0.32
Kutta	2017	19.0	42	42	Math	Week	CC	B	E	NR	<=30 min	Typical	−0.20
Kutta	2017	19.0	42	42	Math	No	CC	B	E	NR	<=30 min	Typical	−0.01
Kutta	2017	19.0	38	38	Math	Week	CC	B	E	NR	<=30 min	Typical	0.10
Kutta	2017	19.0	38	38	Math	No	CC	B	E	NR	<=30 min	Typical	0.16
Lin & Bigenho	2011	22.0	21	21	Mem	No	FM	W	E	no	<=10 min	Typical	−0.75
Lin & Bigenho	2011	22.0	21	21	Mem	No	FM	W	E	Man	<=10 min	Typical	0.07
Lin & Bigenho	2011	22.0	21	21	Mem	No	FM	W	E	Man	<=10 min	Typical	0.15
Luo et al.	2018	21.0	32	34	SS	Min	Mix	B	E	Nat	<=30 min	Typical	−0.58
Luo et al.	2018	21.0	30	30	SS	Min	Mix	B	E	Nat	<=30 min	Typical	0.36
Mitchell & Zheng	2017	19.0	29	29	Math	Min	CC	B	E	Nat	<=20 min	Typical	−0.35
Mitchell & Zheng	2017	19.0	29	29	Math	Min	FM	B	E	Nat	<=20 min	Typical	0.19
Morehead et al.	2019	24.0	32	31	Lec	No	FM	B	E	no	<=20 min	Typical	−0.41
Morehead et al.	2019	23.0	31	31	Lec	No	FM	B	E	no	<=20 min	Typical	−0.32
Morehead et al.	2019	24.0	32	31	Lec	No	CC	B	E	no	<=20 min	Typical	−0.14
Morehead et al.	2019	24.0	32	31	Lec	Days	FM	B	E	no	<=20 min	Typical	−0.09
Morehead et al.	2019	24.0	32	31	Lec	Days	CC	B	E	no	<=20 min	Typical	−0.06
Morehead et al.	2019	23.0	61	61	Lec	Days	CC	B	E	no	<=20 min	Typical	0.18
Morehead et al.	2019	23.0	61	61	Lec	Days	FM	B	E	no	<=20 min	Typical	0.19
Morehead et al.	2019	23.0	31	31	Lec	No	CC	B	E	no	<=20 min	Typical	0.36
Mueller & Oppenheimer	2014	19.0	22	22	Lec	Week	FM	B	E	no	<=10 min	Typical	−0.80
Mueller & Oppenheimer	2014	19.0	28	27	Lec	Week	CC	B	E	no	<=10 min	Typical	−0.80
Mueller & Oppenheimer	2014	19.0	22	22	Lec	Min	CC	B	E	no	<=20 min	Typical	−0.41
Mueller & Oppenheimer	2014	19.0	33	34	Lec	Min	CC	B	Q	no	<=20 min	Typical	−0.34
Mueller & Oppenheimer	2014	19.0	33	34	Lec	Min	FM	B	E	no	<=20 min	Typical	−0.12
Mueller & Oppenheimer	2014	19.0	33	34	Lec	Min	FM	B	Q	no	<=20 min	Typical	−0.06
Mueller & Oppenheimer	2014	19.0	26	28	Lec	Week	FM	B	E	no	<=10 min	Typical	−0.04
Mueller & Oppenheimer	2014	19.0	26	28	Lec	Week	CC	B	E	no	<=10 min	Typical	0.18
Pettit-O'Malley et al.	2017	21.0	32	60	Hum	Days	CC	B	Q	Nat	Days	Typical	−0.39
Quade et al.	1995	21.0	17	56	TC	No	Mix	B	Q	Nat	Weeks	Typical	0.59
Richards et al.	2016	11.0	8	8	OT	Days	NR	W	E	Nat	Semester	Atypical	−0.30
Schoen	2012	20.5	20	18	Mem	Min	FM	B	E	NR	Days	Typical	4.52*
Smoker et al.	2009	19.3	30	30	Mem	Min	FM	B	E	no	<=10 min	Typical	−0.57
Smoker et al.	2009	19.3	30	30	Mem	Min	FM	B	E	no	<=10 min	Typical	−0.55
Smoker et al.	2009	19.3	30	30	Mem	Min	FM	B	E	no	<=10 min	Typical	−0.48
Smoker et al.	2009	19.3	30	30	Mem	Min	FM	B	E	no	<=10 min	Typical	−0.17
Sun & Li	2019	17.5	32	40	TC	Days	FM	B	E	Nat	Days	Typical	0.48
Sun & Li	2019	17.5	32	40	TC	Days	FM	B	E	Nat	Days	Typical	0.57
Sun & Li	2019	17.5	32	40	TC	Days	CC	B	E	Nat	Days	Typical	0.64
Urry et al.	2020	20.0	74	68	Lec	Min	FM	B	E	no	<=20 min	Typical	0.05
Urry et al.	2020	20.0	74	68	Lec	Min	CC	B	E	no	<=20 min	Typical	0.19
Wei et al.	2014	21.9	20	15	TC	No	FM	B	E	no	<=20 min	Typical	−0.64
Wei et al.	2014	21.9	18	27	TC	No	FM	B	E	Man	<=20 min	Typical	0.44

(continued on next page)

Table 1 (*continued*)

Authors	Year	Age	NH	ND	Topic	Interval	Material	Design	Study Type	Distr	Duration	Sample	d
Wood et al.	2012	20.1	21	82	SS	No	Mix	B	E	Man	<=30 min	Typical	−0.35
Wood et al.	2012	20.1	21	42	SS	No	Mix	B	E	no	<=30 min	Typical	−0.01
Zyrowski	2014	16.5	63	63	OT	Days	CC	W	E	Nat	Weeks	Typical	−0.63

Notes. Multiple entries for the same study reflect non-independent effect sizes although the underlying moderator might not appear in the table. NH = Number of participants in the handwritten group; ND = Number of participants in the digital method group; Topic: Mem = Memory, SS = Social Sciences, Lec = General Interest, Hum = Humanities, TC = Technology/Computer, OT = Other; Test Type: Reco = Recognition, Reca = Recall; Interval: Min = Minutes; Material: FM = Factual/ Memory, CC = Comprehension/Conceptual; Design: B = Between, W = Within; Study Type: Q = Quasi-experiment, E = Experiment; Distr = Distraction: Nat = Natural, Man = Manipulated. For all categories, NR = Not reported. For Mueller & Oppenheimer (2014), the retrieved data were based on the Mueller & Oppenheimer (2018) corrigendum.

7. Method

7.1. Study selection

The retrieval of studies was performed through searching PsycINFO, ERIC, SCOPUS, Academic Search Premier, PubMED, SocioINDEX, and Canadian Business and Current Affairs databases in May 2020. The search terms incorporated truncation (or equivalent) in different databases to include all related terms. The terms used were: *("note taking" OR notetaking OR "tak* notes") AND (academic OR learning OR student* OR science OR math* OR reading OR exam* OR course OR class* OR scholastic OR school OR test) AND (achieve* OR perform* OR outcome* OR comprehen* OR marks OR "grade point average" OR GPA OR grades).* Search terms were selected to be as broad as possible. In addition, a proximity search was implemented, reflecting a proximity of two words, by the addition of "N1". The option to "include related terms" was also implemented for all searches, and articles in any language were included. This search resulted in 8,020 non-overlapping hits.

It is important to note that efforts were made to obtain unpublished records to minimize publication bias through a search of "grey literature" on various websites: www.gse.harvard.edu, https://www.eschoolnews.com/, https://www2.gnb.ca/content/gnb/en.html, https://www.

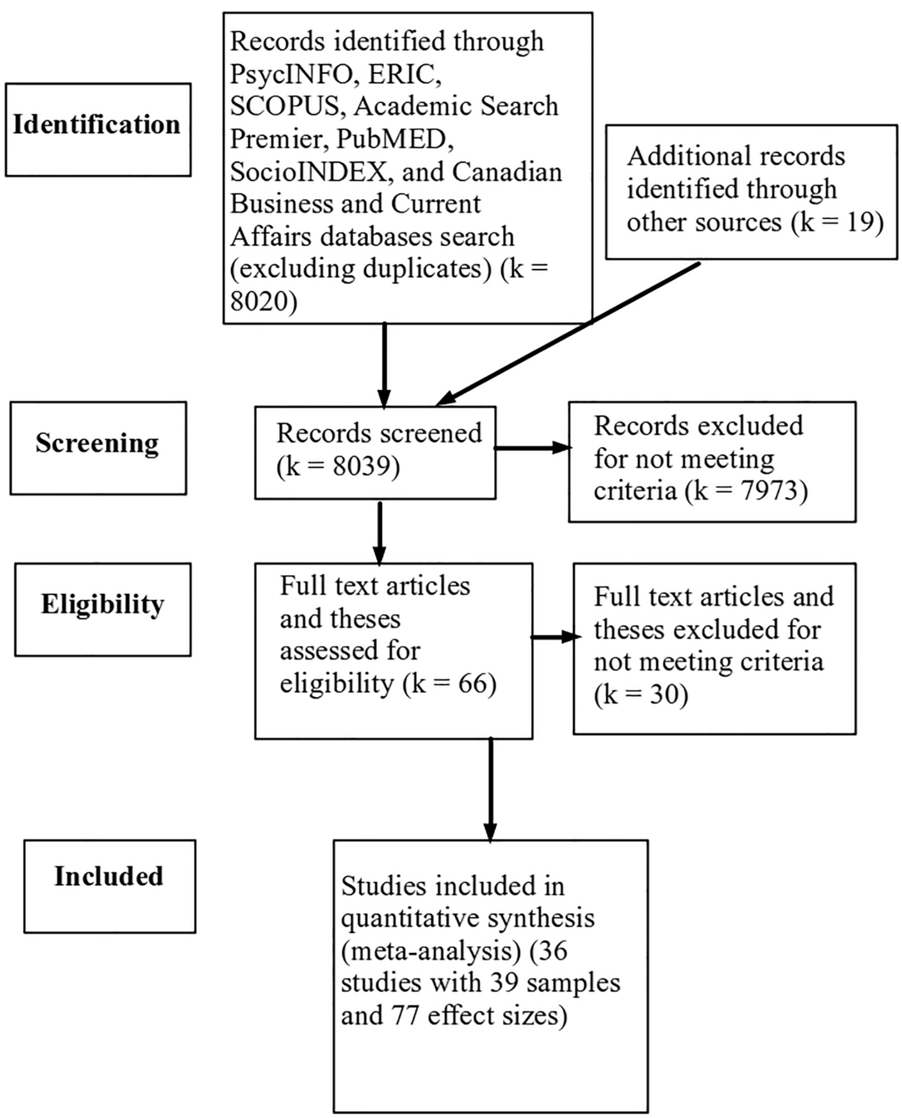

Fig. 1. Flow chart summarizing the data search and inclusion process.

canada.ca/en.html, https://www.google.ca/advanced_search, Psyarxiv, and the Open Science Foundation preprint repository. These efforts resulted in data from 2 unpublished manuscripts. Unpublished theses and dissertations were also included (10 records). In the final sample, 18 effect sizes (23.4%) of the total 77 effect sizes came from unpublished literature. The search process is summarized in Fig. 1.

7.2. Selection criteria

Selection criteria were applied in the process of study retrieval to decide whether a study should be included in the meta-analysis. As a starting point for this process, a paid research associate with meta-analysis training and a clear understanding of the selection criteria described here read the title and abstract for each study to ensure that it fit the basic inclusion criteria. When fit was unclear, the paper itself was consulted to ensure a valid inclusion decision.

The crucial criterion was that studies had to include a comparison of notetaking methods (by hand, on a computer, on a tablet, etc.) and a measure relevant to class performance (course final grade, exam grade, etc.) in any course topic or experimental design at any level of schooling. This criterion underlines the fact that, unlike Allen et al. (2020), we only included studies with objective measures of performance, thereby excluding those with subjective teacher ratings. In the end, only 7 of the 14 studies sampled by Allen et al. warranted inclusion in our meta-analysis.

Data from selected samples of individuals (identified as gifted, with serious cognitive impairments, etc.) were included to ensure we maximized our sample size. However, this component was considered as a moderator in preliminary analyses. Studies were also excluded if any kind of intervention was applied to the notetaking process, although non-intervention control group data from such studies were included. These exclusion criteria distinguish our meta-analysis from the one by Allen et al. (2020), as we aimed to obtain a pure estimate of the comparison between notetaking media. No other exclusion criteria were applied.

Studies ideally had to include relevant data for calculating effect sizes, and the corresponding authors were contacted for studies with information missing published during or after 2010. The year 2010 included a 10-year span from the time when we started data retrieval, and it was selected as a cutoff in the hope it was recent enough to allow a reasonable expectation that authors would still have access to the data. Three authors were contacted in this manner, and one provided the information required to include their study.

Reference lists from retrieved articles were also examined in case some might have been missed in our other searches. Data collection through our searches and communication with authors and the application of the inclusion/exclusion criteria resulted in a sample of 77 effect sizes from 39 samples in 36 different articles. The studies included in the final sample are listed in Table 1.

7.3. Coding of variables

Moderator variables, effect sizes, and other information relevant to each study entered in the analysis were coded in a systematic fashion. The multilevel nature of the design allowed coding of sample-level and measurement-level variables.

At the sample level, age of the sample was coded. It was considered both as a continuous variable and a categorical variable to account for possible non-linearity in age effects. As a categorical variable, age was classified as younger than 13 years, between 13 and 17 inclusive, 18 to 29, 30 to 49, and 50 and above. Although somewhat redundant with age, grade level was coded as elementary, middle school, high school, university/college, and graduate.

Other sample-level moderators were coded to characterize the origin of the sample, although some of them were not published in specific studies. For example, type of school (private vs. public school) was coded as a proxy for socio-economic status (reported for only two of the effect sizes in our sample), although it was relevant only in the context where a course was examined, as opposed to an experimental task. Originally, ethnic origin/racial identity was also considered, but it turned out to be either unreported or reported as a mix in most research. Finally, we identified only two types of samples in the retrieved literature: typical and atypical (a mix of samples with learning or reading disabilities). Year of publication was also an easily available sample-level variable and allowed a consideration of possible cohort effects in the magnitude of effect sizes.

At the measurement level, course or lecture topic (e.g., computer science, math, medicine, research methods, psychology, urban planning, various, word list or other) was considered. For learning objectives, we discovered when coding the studies that most of them did not involve actual class material, but rather they were in a laboratory setting for a one-time study, hence, not a “true” class experience. Therefore, the goal of notetaking was often simply for the participant to remember as much of the material as possible either to be able to recall/recognize it later or to use their memory of it to answer conceptual questions. There were a few exceptions where class grades were at stake ($k = 3$) or the material was relevant to preparation either for a course or a later exam ($k = 3$). Accordingly, we used this rather coarse code (memory, class grade, preparation) to code learning objectives. Note that this last variable was not part of the pre-registration material and was added in response to peer reviewer comments.

As a further measurement-level moderator, whether a distraction was present during the notetaking session was coded as no (typically in controlled experiments), natural (as in a natural course setting), and manipulated. The moderator relevant to the research design (experiment or quasi experiment) was also coded, although it was related closely to the coded factor of whether participants self-selected in their preferred method or if they were assigned randomly to a notetaking method, without regard for their preference.

7.4. Coding validity

We ensured the validity of coding by first preparing a coding sheet where all relevant variables were listed. Two coders independently processed 14 articles accounting for 40 effect sizes (57.1% of total) to assess inter-rater reliability. Coder 1 was a paid research associate with extensive training in the coding process, whereas Coder 2 was an experienced meta-analyst, with over 25 years of experience in meta-analytic coding (Coder 2 also trained Coder 1). Although not all coded variables were used as moderators in the data analysis, we made efforts to achieve a high level of accuracy in coding. Specifically, there were 27 variables involved in the evaluation of coding validity: year of publication, sample id (crucial for multilevel meta-analysis), publication status, school grade, mean age, school type (private, public, etc.), ethnic/racial origin, socio-economic status, country of testing, number of participants in longhand group, number of participants in digital group, method (experiment or quasi-experiment), measure of performance, whether studying was allowed, delay between learning and testing, delivery method (auditory, visual, etc.), testing method (paper, digital, etc.), scale of measurement (accuracy, grades, exam, etc.), test type (recall, recognition), material (conceptual or factual), design (between- or within-subject), whether distractions were present, material topic, duration of material, whether participants self-selected into groups, sample (typical, atypical), statistic used to compute the effect sizes, and effect size.

This coding process produced an overall inter-rater reliability of 98.6% (Kappa = 0.97, 95% $CI = 0.96–0.98$). Furthermore, among coded variables where at least one disagreement occurred, we found 6 disagreements on the 40 double-coded effect size estimates (Kappa = 0.70 (95% $CI = 0.54–0.86$). In the context of study relevance to the inclusion/exclusion criteria (based on the final theses and articles assessed for eligibility; see Fig. 1), there were 5 disagreements on 66 hits (Kappa =

0.85, 95% $CI = 0.76–0.94$). For sample ID, number of participants in the digital group, performance measure, material, and publication status, only 1 disagreement occurred over 40 entries, for Kappa = 0.95 (95% $CI = 0.88 – 1.0$) in each case. For the remaining variables, Kappa would be 1.0 as no disagreements occurred. This high level of inter-rater reliability suggests that the coding sheet was generally valid and reflected a straightforward process. For the remainder of the coding process, Coder 1 coded the remaining papers independently. As a further step to ensure data validity, the three authors of this paper (including the two original coders) held regular meetings to discuss any coding issues or problematic papers. As a result of this multi-step coding process, we are confident the final data provided a valid reflection of the sampled literature.

7.5. Measure of effect size

As we were dealing generally with designs reflecting the comparison of two groups, the standardized mean difference between the performance of each notetaking method group or Cohen's *d* (Cohen, 1988) was the measure of effect size. In our computations, a positive value reflected an advantage for digital notetaking. As means and standard deviations were available either from the articles or after contacting authors for 64 of 77 effect sizes (83.1%), we were able to calculate these effect sizes with the formula presented by Cohen (1988). In most other cases, an *F, t,* chi square, correlation, or a *p* value was available, allowing use of the formula presented by Lipsey and Wilson (2001). It is important to note that, although 68 of 77 effect sizes (88.3%) were derived from a between-subject design, the remainder came from a within-subject design. Computation of Cohen's *d* in such a setting typically requires the standard deviation of the difference as well as the correlation between the two scores. However, the first value was not reported in any of the retrieved papers, and the second one was reported in only one study. Accordingly, as a rudimentary way to adjust these effect sizes, we used 0.70 as a conservative estimate of correlation and multiplied the Cohen's *d* obtained from the between-subjects equation by 2*sqrt(1-*r*) (see Borenstein et al., 2009, for more details).

All effect sizes were computed by means of the calculator created by David Wilson (http://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php). Furthermore, the Hedges and Becker (1986) small samples correction was applied to the effect sizes (nominally, Hedges' *g*).

7.6. Data analysis

Considering that the sampled studies had multiple measures of performance or non-independent effect sizes related to our moderators, we implemented a multilevel model in our data analysis (Raudenbush & Bryk, 2002; Van den Noortgate et al., 2013, 2015). We also computed robust standard errors in testing moderators to reduce the risk of Type 1 error while considering within-study sampling variability. Considering that we obtained a relatively small sample of 77 effect sizes, all analyses were conducted for the overall sample based on the notion that breaking down into smaller groupings (e.g., by lecture topic) would produce small clusters and considerably reduce statistical power.

We also conducted publications bias analyses on two different fronts. First, we compared published and unpublished work, with the expectation that if a publication bias existed, it would show smaller effect sizes for the latter compared to the former. We also presented a funnel plot (effect sizes on the X axis, standard error on the Y axis) and conducted the analysis proposed by Egger, Davey Smith, Schneider, and Minder (1997). As proposed by these authors, evidence for a publication bias would be shown if regression of the standard normal deviate for the effect size on the inverse of the standard error showed an intercept or slope significantly different from zero, with $p < .10$.

All meta-analytic results were computed by means of the meta*for* package in the R statistical software (Viechtbauer, 2010), with effect sizes treated as random effects and moderators treated as fixed effects. The estimates obtained in this approach reflect precision-weighted estimates of effect sizes as in other approaches to meta-analysis (Raudenbush & Bryk, 2002).

Categorical independent variables were dummy coded with the use of "as.factor()" in R, producing $c - 1$ dichotomous vectors (*c* represents the number of categories). Continuous moderators were mean-centered. All significance testing for moderators was based on the omnibus test obtained through examination of robust estimates of standard errors computed with the *robust* procedure available in meta*for*. In addition, we implemented the correction for small samples built in the *robust* procedure. Only moderators that contributed significantly to variability with $p < .05$ are presented.

8. Results

8.1. Preliminary analyses

Considering that our sample included within-subject designs for which we computed only an approximation of the appropriate effect size, we believed it was important to determine whether this factor unduly affected the results. Accordingly, we conducted a moderator analysis with design (between or within) to decide if such studies fit within our population of effects. Results of this analysis showed that design accounted for significant variability, $F(1, 37) = 5.52$, $p = .024$. This reflected the finding that within-subject designs produced a mean effect size of −0.29 (95% CI: −0.47, −0.11) that was significantly smaller than zero, whereas the mean effect size of 0.01 (95% CI: −0.17, 0.20) for between-subjects design was not significant. This suggests the initial conclusion that within-subject designs might need to be excluded from further analyses, as they do not seem to belong in this population of effect sizes.

As a further preliminary analysis, we also considered the comparison of typical and atypical samples in a moderator analysis. Results for this analysis showed that sample type accounted for significant variability, $F(1, 37) = 9.40$, $p = .004$. This reflected the finding that atypical samples had a mean effect size of −0.28 (95% CI: −0.28, −0.27) that was significantly smaller than zero, whereas the mean effect size of 0.01 (95% CI: −0.19, 0.16) for typical samples was not significant. Again, this would suggest it is prudent to exclude atypical samples because they might not belong in our population of effect sizes. However, sample and design were confounded in that all atypical samples were from within-subject designs. We, therefore, re-ran the analysis on design as a moderator with the atypical samples excluded, and this showed that it no longer accounted for significant variability, $F(1, 35) = 3.67$, $p = .064$. Accordingly, we conducted all further analyses with atypical samples excluded but with the remaining within-subject designs ($k = 5$) included.

As a final preliminary analysis, we defined outliers as those that deviate from the overall mean by >3.29, which Tabachnick and Fidell (2007) suggest as reasonable because it would exclude scores that have a probability smaller than 0.001 of belonging in the sample. This approach produced three outliers from two samples. However, conducting the analyses with and without these outliers did not affect the results in terms of significance of the mean effect sizes or in determining significant moderators. Accordingly, they were preserved in all data analyses, although they are starred in Table 1 for interested readers. Overall, data analyses proceeded with a final sample of 73 effect sizes from 37 samples, reflecting data from 3,120 participants.

8.2. Overall results

The overall analysis revealed a mean effect size of −0.008 (95% *CI*: −0.18, 0.16) that was not significantly different from zero. Therefore, overall results suggest no effect of notetaking approach (longhand or digital) in the overall sample. However, we also found that the effect sizes were heterogeneous, $Q(72) = 1351.69$, $p < .001$, $I^2 = 80.9\%$. Thus, the overall effect size is not representative of the state of affairs in our data. This also indicates that moderator analysis is warranted to identify factors that might account for this variability.

8.3. Moderator analyses

Moderator analyses proceeded with a systematic examination of all the moderators described in the method section. Two moderators relevant to assessment proved significant. Topic was the only pre-specified moderator that accounted for significant variability in effect sizes, $F(9, 27) = 749.59$, $p < .001$. Relevant summary data presented in Table 2 reveal three categories where effect sizes were significantly different from zero: urban planning and not reported material showed an advantage for handwritten notes, whereas computer science material produced an advantage for digital devices. Although no other pre-specified moderators accounted for significant variability in effect sizes (all p's > 0.52), learning objectives emerged as significant, $F(2, 34) = 9.98$, $p < .001$, with summary values also presented in Table 2. In this case, results showed a significant advantage for handwritten notetaking when class grades were relevant to the objectives, whereas the confidence intervals for memory or preparation for a course/exam included zero. To account for further variability in the effect sizes, we conducted exploratory moderator analyses with additional variables that we had not pre-specified.

In particular, we considered whether study time was provided and the amount of time between the notetaking session and performance as potential moderators, as we surmised that study time and a shorter time interval between learning and testing might reduce the effect of notetaking method. Duration of the material (ranging from less than ten minutes to weeks, when a whole academic term was included in the measure of performance) was an additional moderator, although we had no specific prediction for this variable. Delivery method (written, auditory, mix) was considered based on the notion that the mix of written and auditory presentation found in a typical lecture might produce a significant memory load, making this situation propitious for a larger effect of notetaking method.

Results of these exploratory analyses showed that, in this set of moderators, only material duration, $F(6, 30) = 14.89$, $p < .001$, accounted for significant variability in effect sizes (all p's > 0.11 for the other exploratory moderators). Relevant estimates, presented in Table 2, show that only material with a duration between 30 and 60 min produced a mean effect size that differs from zero. Again, it was negative, reflecting an advantage for handwritten over digital notetaking in that time interval.

As a further exploratory analysis, we considered the possibility that testing method and material type might interact with each other. Clearly, the possibility that recall or conceptual questions might produce a significant advantage for handwritten notetaking did not materialize in the relevant main effects. However, the possibility remains that these two variables might combine to produce large effects for the recall of conceptual material presumably due to the depth of processing required to achieve optimal performance under such difficult conditions. Unfortunately, pursuit of these analyses was undermined by the nature of the retrieved data. Specifically, as seen in Table 3, many categories had few effect sizes, with three cells containing only one. Although this makes a test of significance of the interaction invalid, the effect sizes reported in the table suggest large fluctuations as a function of each variable combination. Even then, some of the estimates cannot be interpreted because they were derived from a single effect size. To somewhat circumvent this problem, we only considered the clear-cut conditions where test type and material type were clearly defined (recall/factual, recall/conceptual, recognition/factual, and recognition/conceptual). With only these four cells, the smallest cell size was three, and it was possible to test the interaction. Results of this analysis showed that the interaction was not significant under these circumstances ($p = .067$).

8.4. Publication bias

As a first step, we examined whether publication status (published versus unpublished) affected the magnitude of effect sizes. Results of this analysis showed that publication status was not a significant moderator of effect sizes, $F(1, 35) = 2.25$, $p = .14$. This analysis therefore does not support the presence of a publication bias. As shown in the funnel plot of precision as function of effect size in Fig. 2, effect sizes were symmetric, further supporting the notion that there is no evident publication bias in the data. In addition, as recommended by Rodgers and Pustejovsky (2020), we used a multilevel extension of the Egger et al. test (Egger MLMA test) with the modified precision estimate proposed by these authors as a covariate and effect size as the outcome variable. A significant slope in this method indicates the presence of funnel plot asymmetry and is interpreted as evidence of a publication bias. To maximize power, we still applied the recommendation by Egger et al. to use a significance level of 0.10. Results of this analysis showed that the slope ($b1 = 1.13$, 90% CI: −1.44, 3.69) was not significant. It therefore appears that based on the three sources we used here, there is no

Table 2
Results for material content and material duration as moderators.

Moderator	Categories	k	g (95% CI)
Topic	Computer science	4	0.57 (0.55, 0.59)*
	Math	9	0.02 (−0.17, 0.22)
	Medicine	5	0.10 (−0.80, 0.99)
	Research method	3	0.13 (−0.59, 0.85)
	Not reported	5	−0.58 (−0.60, −0.57)*
	Other	5	−0.06 (−0.34, 0.23)
	Psychology	8	0.34 (−0.37, 1.04)
	Urbanism	3	−0.52 (−0.73, −0.31)*
	Various	19	−0.19 (−0.46, 0.07)
	Word list	9	−0.19 (−0.46, 0.09)
Learning Objectives	Class grades	3	−0.38 (−0.44, −0.33)*
	Memory	66	0.02 (−0.18, 0.23)
	Preparation	4	0.01 (−0.31, 0.33)
Material Duration	<10 min	18	−0.05 (−0.54, 0.45)
	10–20 min	25	−0.10 (−0.27, 0.07)
	20–30 min	13	0.02 (−0.18, 0.21)
	30–60 min	5	−0.58 (−0.60, −0.57)*
	Days	6	0.49 (−0.87, 1.84)
	Weeks	2	−0.07 (−1.03, 0.89)
	Full semester	4	0.13 (−0.33, 0.59)

Note. The table presents the number of effect sizes (k) and the mean weighted g for each moderator category with the 95% confidence interval (CI) in parentheses.
*The mean weighted effect size is significantly different from zero with $p < .05$ when the 95% CI for g does not include zero.

Table 3
Results relevant to the interaction between test type and material type.

Test type	Material type	k	g (95% CI)
Recall	Factual	21	−0.13 (−0.33, 0.06)
Recognition	Factual	8	−0.06 (−0.36, 0.25)
Both	Factual	3	1.45 (−1.06, 3.96)
Recall	Conceptual	17	0.004 (−0.27, 0.27)
Recognition	Conceptual	3	−0.33 (−0.77, 0.11)
Both	Conceptual	1	**1.39 (−0.64, 3.42)**
Recall	Mix	1	**−0.38 (−0.53, −0.23)**
Recognition	Mix	11	−0.05 (−0.25, 0.16)
Both	Mix	1	**0.59 (0.59, 0.59)**
Recall	Not reported	2	−0.24 (−0.96, 0.49)
Recognition	Not reported	3	0.07 (−0.94, 1.08)
Both	Not reported	2	−0.43 (−0.64, −0.22)*

Note. The table presents the number of effect sizes (k) and the mean weighted g for each moderator category with the 95% confidence interval (CI) in parentheses. *The mean weighted effect size is significantly different from zero with $p < .05$ when the 95% CI for g does not include zero. Bolded estimates cannot be interpreted because they were derived from a single effect size.

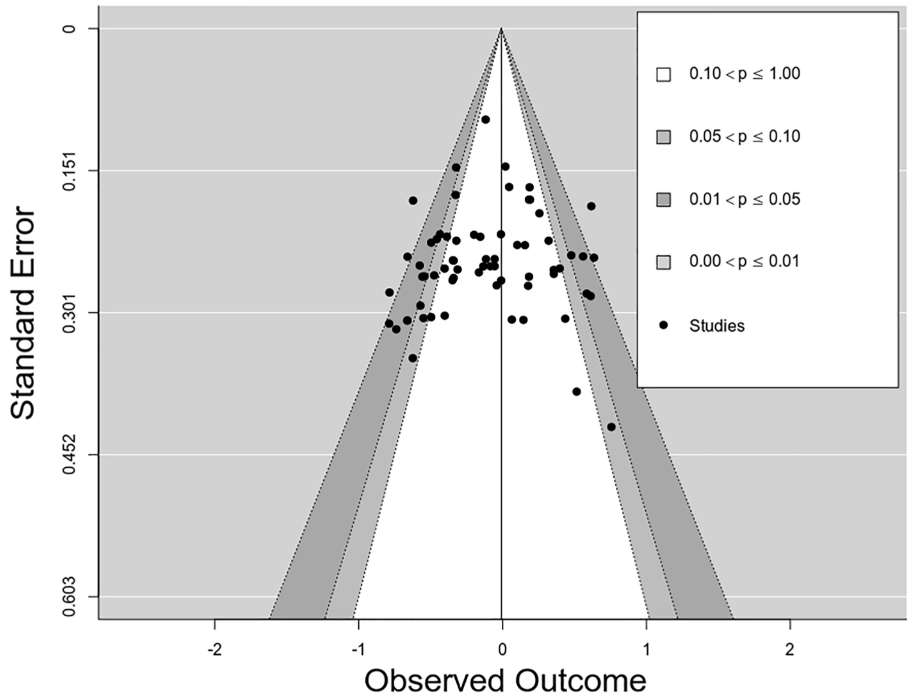

Fig. 2. Contour enhanced funnel presenting precision (standard error) as a function of Hedge's g (observed outcome) for the overall sample.

evidence of a publication bias in our data.

9. Discussion

The purpose of the present meta-analysis was to fill a gap in the literature by developing a more definitive understanding of whether notetaking method (longhand writing or digital) affected academic performance. We pursued this goal by retrieving literature from as many sources as possible (published and unpublished) and by integrating relevant effect sizes into a multilevel meta-analysis that also took into account correlated sampling errors by means of robust variance estimation. In assessing whether the notetaking method itself made a difference, we excluded additional distractions that might contaminate the use of some digital methods in the classroom.

9.1. Overall results

Overall, when distractions are not a factor or are equated across methods, our findings from a comprehensive set of the relevant literature demonstrated that notetaking method makes no difference on performance. This is in contrast to what was reported in Mueller and Oppenheimer (2014), the most referenced single paper to date on this research question. These authors found that the use of handwritten notes produced better performance than digital notetaking when the content was conceptual rather than factual. They interpreted their results as supporting the notion that longhand notetaking promotes deeper processing than taking notes on a laptop, in keeping with a level of processing interpretation. However, in the current study, even when we considered content (conceptual or factual), there was no effect on the magnitude or direction of the effect sizes. Thus, although individual studies can produce results supporting one view over another, the cumulative evidence suggests that notetaking method makes no difference on performance.

It is important to consider the possibility that some readers might view the present overall results as amounting to support for the null hypothesis. However, in the context that most results should be subjected to "meta-analytic thinking" (Thompson, 2002; see also Ritter, 2015 for views relevant to educational technology research), it would be inappropriate to deal with the present results otherwise. Briefly, meta-analytic thinking refers to the need to consider a result for a given study in the context of other existing research, the observed confidence intervals, and relevant meta-analytic findings. For the present study, meta-analytic thinking refers to the fact that we were mostly concerned with establishing the magnitude of the effect and its probable range in terms of confidence interval. Although important, whether the true value of the effect size is significant is not central to its interpretation. Based on the premise that we have as comprehensive a sample as possible and used appropriate methods, our results show that when we compare longhand and digital notetaking methods equated on distractions, the true effect is located in a range that includes zero. In practice, this means that these two methods are not distinguishable in a comprehensive set of the literature. Accordingly, the present results should be seen as moving research forward, despite the lack of significance for the overall effect. The interpretation is straightforward, and it reflects the question that guided the purpose of the present meta-analysis. Essentially, based on the available literature, when longhand and digital methods are treated as pure notetaking tools (i.e., equated on distractions), then the effect of this manipulation on performance is not distinguishable from zero.

9.2. Variability and moderator effects

Despite a clear general conclusion, it is very important to point out that the effect sizes in the overall sample were quite variable. In fact, in interpreting the I^2 that we obtained, results showed that 80.9% of the total variance was accounted for by heterogeneity in the overall sample of effect sizes (i.e., variance in the true effects), whereas the remaining 19.1% reflects random sampling variance (i.e., chance). This amount of variability can be categorized as substantial (see https://handbook-5-1.cochrane.org/chapter_9/9_5_2_identifying_and_measuring_heterogeneity.htm) and makes it rather difficult to interpret the overall effect as it cannot be considered representative of the state of affairs in the sample (Borenstein et al., 2009). In view of this variability, it made sense to conduct moderator analyses to explain at least some of it.

Even in that context, results of the moderator analyses proved disappointing, with only one of our pre-specified moderators (topic) and two exploratory moderators (learning objectives, material duration) accounting for significant heterogeneity among the effect sizes. Nevertheless, interpretations for these significant moderators offer some interesting speculations that might guide future work while also offering potential for pedagogical advice. Specifically, as shown in Table 2, studies where the topic was on urban planning produced a significant advantage for longhand notetaking, whereas computer science material resulted in improved performance for digital notetaking. Considering that rote memorization is often a preferred approach in learning and assessment for some of the sciences (Edmondson & Novak, 1993; Pabuccu & Erduran, 2017), it is possible that urban planning courses, relevant to the field of Humanities, might involve more critical thinking (Dumitru, 2019) and therefore may emphasize deeper level of processing and higher overall cognitive load. This would fit with the notion that longhand notetaking can preserve this deeper level of processing. In contrast, the computer science studies that entered the analysis relied on participants who had an existing level of comfort with digital devices (at least with computers). For example, Quade et al. (1995; contributing one effect sizes to the computer science category) tested only computer science majors and minors. Similarly, Sun and Li (2019; accounting for the three computer science effect sizes here) found a significant effect of notetaking method in favor of the digital approach only for "excellent students" in their samples (compared to mid-level and low-performing students). Therefore, findings with computer science material in our sample might reflect existing level of comfort with computers. Still, it might be worthwhile examining more closely the role of comfort with computers in explaining differences in performance as a function of notetaking method. Of course, we can only speculate on possible causes for the effect of topic on the direction and magnitude of the observed effect sizes considering that only three effect sizes each entered the computer science and urban planning categories. In addition, the "not reported" category ($k = 2$) was the other case where a significant longhand notetaking emerged but the fact that the topic is undefined makes it difficult to draw solid conclusions on findings in that category. Nevertheless, our findings suggest that researchers should consider closely the topic of the lecture and individual differences in comfort with computer usage.

Although coarse, our coding of learning objectives produced the interesting finding that only class grades produced a significant advantage for handwritten notetaking. When compared with the lack of effect for studies where memory for the items (for factual or conceptual purpose) or where preparation for a later event is involved, the significant effect size for class grades suggests that the notetaking method might matter only when there is something personal at stake in the evaluation. This raises the possibility that higher depth of processing favors handwritten notetaking in such a case. However, here as well, we can only speculate considering that we could only retrieve three effect sizes obtained with class grades. This limits our ability to draw definite conclusions, but it suggests strongly that future work using high-stakes situations is required in this domain.

For material duration, we found that material with a duration between 30 and 60 min produced a significant advantage for the longhand method, whereas we found a medium effect size suggesting some (non-significant) improvements with digital notetaking when material was spread across days. We could speculate that the 30 to 60 min material duration would reflect a typical lecture that somehow favored performance after longhand notetaking. However, it is difficult to interpret why that duration should be particularly meaningful in terms of the effect of notetaking method on performance, especially when we expected larger effects for longer duration material.

Based on conceptual grounds, we also conducted an exploratory analysis considering a possible interaction between material type (factual, conceptual) and test type (recall, recognition). A clear examination of this interaction was hindered by cells that included only one effect size. Limiting the analysis to clear-cut cases, thereby excluding conditions where a mix of material type or a combination of both test types was used, did not produce a significant interaction. However, even then, only three effect sizes reflected recognition for conceptual material, thereby limiting the precision of these estimates and the underlying statistical power. Nevertheless, findings in Table 3 show that the largest effects (regardless of direction) were obtained when assessment included both recall and recognition test types in a single measure are intriguing. This suggests that more research that systematically investigates the potential interaction between test type and material type is needed.

Overall, our results leave us with the conclusion that, taken as a whole, the existing literature suggests that the notetaking method has little influence on performance when distractions are equated across methods. The corollary to this conclusion is that findings of better performance under digital compared to longhand notetaking by Allen et al. (2020) are likely fully accounted for by the distraction component that was left uncontrolled in their sample for digital methods. Specifically, email, irrelevant internet browsing, and social media offer many avenues for distractions when using digital means of notetaking (Dontre, 2020). In an ideal world, one solution to eliminate irrelevant distractions for digital notetakers would be to make the internet or the phone signal unavailable during class. However, this is not practical in most cases for technical or safety reasons. Another way to eliminate (or, at least, reduce) distractions from digital devices takes into account the fact that, ultimately, the decision rests with the students on whether they will indulge in what has been called cyberslacking (Rana, Slade, Kitching, & Dwivedi, 2019). Specifically, Rana et al. showed that a large set of factors (i.e., lack of attention, apathy towards course material, distraction by others, attitude toward cyberslacking, subjective norm, perceived behavioral control, perceived threat, and escapism) were significant predictors of intention to indulge in cyberslacking. In reality, factors such as lack of attention, apathy toward course material, and escapism could be mitigated by providing engaging lecture material to students. In addition, setting clear rules for class behavior would set the perceived norm against (rather than in favor) of cyberslacking. Of course, these are only speculations, but they emphasize the need for instructors to set clear rules about appropriate classroom behavior and maximize student engagement to reduce the influence of distraction on classroom behavior (including notetaking efficiency).

In attempting to explain our results, it is also important to keep in mind that there is considerable variability in the effect sizes that cannot be explained by the moderators we had pre-specified and those we explored as a further attempt to account for heterogeneity among the effect sizes. This points to the possibility that many of the observed effects are affected by sample specific or idiosyncratic components of the various studies. It also suggests the presence of noteworthy limitations in the present meta-analysis, which we will now consider as well as future directions to address them.

9.3. Limitations & future directions

When considering limitations, the null effect of distractions as a moderator is an interesting finding that requires some discussion. Our goal was to get a pure estimate of notetaking method when distractions were removed or equated across conditions, and our overall results suggest that the two notetaking methods do not differ from each other when distractions are equated across them. However, the lack of moderator effects for the presence of distractions is puzzling as, based on the Allen et al. (2020) meta-analysis, one might expect that the presence of distractions would favor longhand notetaking. To better understand our results, it is important to remember that to remove distraction as a confound, unlike Allen et al., we did not include studies where the distraction was only applied to the digital task. In addition, we were only able to identify four studies (five effect sizes) where distraction was manipulated, and these studies showed disparate effect sizes (see Table 1), with two negative effects, two positive effects, and one near zero. In total, more work should consider manipulation of distractions during notetaking under both longhand and digital conditions to document the relative influence of this factor on performance.

At a basic level, the purpose of the present meta-analysis relied on comparing conditions where longhand and digital notetaking conditions were equated on distractions. However, it is important to keep in mind that none of the research that entered our meta-analysis accounted for and partitioned out variance in outcomes explained by the distraction imposed by digital notetaking over and above distractions shared across modalities. One way to ensure such control experimentally would be to implement catch trials of some sort to ensure that attention stays on task. However, none of the quasi-experiments that entered the meta-analysis could account empirically for these sources of variations. In the case of experiments in our meta-analysis, they relied on a basic form of probabilistic control, assuming that random assignment would produce equivalent groups on all but the manipulated variables (i.e., notetaking method). In reality, this lack of partitioning of sources of variability in distraction between groups does not change our conclusions because there was some form of control for distraction at least for the experiments that entered our database. Considering that the type of design (experiment or quasi-experiment) was not a significant moderator, base differences in initial level of distraction between groups are unlikely to have affected our results. In the context of any research endeavor, group differences in base distraction propensity could be problematic because in any research that we conduct, there is never a guarantee that a given participant will stay on task. In the classroom context, recent work conducted by Kane et al. (2021) suggested that task-unrelated thoughts (TUT) can be predicted by factors such as seating location (increasing amount of TUT from the front to the back), initial interest in a topic, multi-tasking habits, and initial propensity for mind wandering and boredom. Thus, although the literature we sampled did not address possible individual differences in distraction propensity, the research by Kane et al. suggests possible ways to verify group equivalence in future research comparing notetaking methods in the classroom.

A related issue concerns the argument that distractions are endemic to digital devices and simply cannot be removed. However, such an argument would disregard the possibility to minimize or eliminate distractions found with digital devices through instructions, reward systems, or by scrambling internet signals, to name a few options. Furthermore, even if one does not have the ability to remove distractions found with digital devices, there are many encouraging sources of evidence suggesting that distractions that come from other sources can be equally problematic. For starters, the aforementioned research by Kane et al. (2021) suggests that many factors other than the selected notetaking device can cause distraction. In addition, Tesch et al. (2011) reported that students ranked peers talking and a hard-to-understand professor as more distracting than laptop use. Similarly, Blasiman et al. (2018) found that students claimed having a conversation was more distracting than playing a video game when in class. In fact, according to Zeamer and Fox Tree (2013), any sounds incongruent with the lecture can have a negative impact on recall of information. Therefore, even those who use handwritten notetaking could be subject to distractions of this type. In fact, mind wandering can happen regardless of notetaking method, with estimates ranging from 10 to 40% of the time spent on mind-wandering (Unsworth & McMillan, 2017; Wammes, 2016). As a complement to the mind wandering data, Benzimra (2016) and Kay et al. (2017) suggested that students are on task 80% of the time when using digital devices, whereas Kane et al. (2021) reported an average rate of task-unrelated thoughts of 24%. Therefore, it would be an exaggeration to state that distractions are endemic only to digital devices. In reality, students can be distracted regardless of their notetaking method, but, thankfully, a large majority of them are likely to be on task at any given point during a lecture. In this context, experimental research that promotes distractions under the digital-device conditions when comparing notetaking methods might overestimate the influence of this factor compared to natural conditions.

Our focus on performance measures rather than on other potential benefits of the notetaking method might be seen by some as a limitation of our meta-analysis because, in that respect, it has a relatively narrow scope in terms understanding positive or negative consequences of the method. For example, our focus on performance does not account for the influence of method on how digital notetaking could distract other students even when they use longhand notetaking, how it might affect student's engagement or motivation, or how it might influence students' self-reports on learning outcomes. However, our reasoning was that performance (i.e., grades) is the most tangible component to be considered. Therefore, the influence of notetaking method on the other aforementioned outcomes still remains to be examined comprehensively.

The focus on performance measures was also affected by the limits of the performance measures themselves. Specifically, as mentioned previously, very few of the studies we found included performance measures reflective of "true class" performance. Instead, there were many relatively simple memory tasks either of a factual or conceptual nature. This makes it quite clear that future research examining true class performance measures is required.

An additional limitation inherent to the generally simple performance measures that have been examined in the existing literature is that results in these studies can only reflect the direct effects of notetaking method on memory, not the interconnected processes that follow from efficient notetaking as one prepares to write an exam for example. From this perspective, notetaking is a necessary but insufficient component that does not capture the additional aspects of learning in a class context such as the restudy of notes and the potential enactment of generative strategies such as retrieval practice, self-explanation, concept mapping, rehearsal, and comprehension activities that would promote learning (Blasiman et al., 2017; Chew, 2021; Rawson et al., 2018). In addition, notetaking in the context of a classroom would be guided by learning objectives that might define focal points in the notes. The existing literature generally does not capture these important aspects and filling this gap should be viewed as a priority for future work.

Effects may change with shifting ways of pedagogical methods (e.g., online/virtual instruction, development of notetaking applications). The constant shift means that it is important to re-examine roles of notetaking. Although performance is an important criterion, it is likely that changes to teaching and learning approaches will occur due to reasons that are not necessarily relevant to general pedagogical standards (e.g., COVID-19 pandemic, administrative mandates, increased globalization, modernization of technology). As such, re-examination will be important simply due to the reality of changing situations.

The finding that we identified few significant moderators could be seen as a limitation of our meta-analysis. However, in reality, it points to limitations in the existing pool of studies. Specifically, few of the moderators that have been covered in the literature appear to affect the magnitude of the effect, leaving us with a sense that the effect of notetaking method is quite variable and sample specific. More work is required to identify potential moderators that account for its variability. More importantly, the existing literature also exhibits issues with the samples used in attempts to establish robust effects. In particular, the literature we retrieved for our meta-analysis shows that much of the research is conducted with small samples (median = 65.00; range = 16 to 430, skew = 2.91), with 66 of 77 effect sizes (85.7%) drawn from university or college students. These are selective samples that are not representative of the population and limit our ability to generalize beyond university students. In fact, the largest sample we found (total N = 430, from Artz et al., 2020) also relied on data from university students. This situation strongly suggests that more research with large representative samples and with children old enough to use digital notetaking is required to fully elucidate the question of whether notetaking method affects learning outcomes. From this perspective, the present meta-analysis should not be seen as the final word on this question but rather as a source of inspiration for more generalizable and better controlled work with more diverse samples.

Finally, a meta-analysis that relies primarily on published work always has to be concerned with the possibility that a publication bias might affect its results. In our case, however, our sample involved a sizeable segment (23.6%) of unpublished studies. Effect sizes for unpublished work did not differ from those obtained with published work. In addition, effect sizes showed a fairly symmetrical distribution (see Fig. 2), and formal statistical tests showed no evidence of a publication bias. Accordingly, we should be confident in stating that our research was influenced minimally by such a bias.

10. Conclusions

The purpose of the meta-analysis presented here was to quantify overall effects of notetaking method, defined as longhand or digital, in various learning situations and to identify potential moderators of these effects. We began this meta-analysis fully expecting to find evidence favoring the use of longhand notetaking over digital notetaking for optimal performance. In fact, anecdotally, two of us who are instructors often suggested to students at the beginning of each semester that longhand notetaking should be favored over digital notetaking. Results showed a null effect of this factor from our comprehensive sample. This suggests that, in general, instructors should be cautious in recommending that one approach is broadly better than another. In reality, our recommendations to instructors would be: If you think you can remove distractions on digital devices one way or another, then, in general, the notetaking method will make no difference. If you cannot control distractions, you might want to discourage the use of digital notetaking.

Mechanistic constraints could be another reason one might have to discourage the use of digital devices in the classroom, based on the notion that longhand notetaking might afford more flexibility as well as the possible use of generative drawings, for example. However, software designers seem to be aware of this potential limitation and have been developing notetaking applications that attempt to resemble the process used for handwritten notes, even providing an ability to include drawings and diagrams (e.g., Microsoft OneNote, Notability). Still, most applications cannot yet compete with the flexibility afforded by handwritten notetaking. Nevertheless, if students can find an application that meets their needs, mechanistic constraints should not be a reason to discourage the use of digital devices for notetaking.

Although our moderator analysis did not highlight specific variables, we are still convinced that the optimal notetaking approach depends on a multitude of factors, only some of which have been examined adequately in the literature. Accordingly, the hunt is still on for factors relevant to the effect of notetaking methods. More research investigating new moderators with large representative samples of all reasonable ages is needed for a more thorough examination of this question.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research was funded by the Teaching and Learning Priority Funds from the University of New Brunswick to S. Ronis and D. Voyer, and by a University of New Brunswick University Research Scholar Award to D. Voyer.

References

Starred (*) references were included in the meta-analysis.

Allen, M., LeFebvre, L., LeFebvre, L., & Bourhis, J. (2020). Is the pencil mightier than the keyboard? A meta-analysis comparing the method of notetaking outcomes. *Southern Communication Journal, 85*(3), 143–154. https://doi.org/10.1080/1041794X.2020.1764613

*Aragón-Mendizábal, E., Delgado-Casas, C., Navarro-Guzmán, J.-I., Menacho-Jiménez, I., & Romero-Oliva, M.-F. (2016). A comparative study of handwriting and computer typing in note-taking by university students. *Comunicar, 24*(48), 101–107. https://doi.org/10.3916/C48-2016-10

*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*, 103–115. https://doi.org/10.1080/00220485.2020.1731386

Barrett, M. E., Swan, A. B., Mamikonian, A., Ghajoyan, I., Kramarova, O., & Youmans, R. J. (2014). Technology in note taking and assessment: The effects of congruence on student performance. *International Journal of Instruction, 7*, 49–58.

Beck, K. M. (2014). Note taking effectiveness in the modern classroom. *The Compass, 1*, Article 9.

*Bennett, J. (2018). *Determining the effectiveness of a note-taking intervention in a 1-to-1 iPad mathematics classroom* [Unpublished Doctoral dissertation]. Robert Morris University.

Benzimra, D. (2016). *An examination of laptop-based off-task behaviours in secondary school classrooms* Unpublished Masters thesis, Department of Education, University of Ontario Institute of Technology. Retrieved from http://hdl.handle.net/10155/665.

Blasiman, R. N., Dunlosky, J., & Rawson, K. A. (2017). The what, how much, and when of study strategies: Comparing intended versus actual study behaviour. *Memory, 25*(6), 784–792. https://doi.org/10.1080/09658211.2016.1221974

Blasiman, R. N., Larabee, D., & Fabry, D. (2018). Distracted students: A comparison of multiple types of distractions on learning in online lectures. *Scholarship of Teaching and Learning in Psychology, 4*(4), 222 230. https://doi.org/10.1037/stl0000122

**Blankenship, M. W. (2016). *Comparing note taking and test performance in methods and modes of note taking conditions* [Unpublished doctoral dissertation]. Ball State University.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. (2009). *Introduction to* meta-*analysis.* Chichester, UK: Wiley.

Bui, D. C., & Myerson, J. (2014). The role of working memory abilities in lecture note-taking. *Learning and Individual Differences, 33*, 12–22. https://doi.org/10.1016/j.lindif.2014.05.002

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

Cabeza, R., Kapur, S., Craik, F. I. M., McIntosh, A. R., Houle, S., & Tulving, E. (1997). Functional neuroanatomy of recall and recognition: A PET study of episodic memory. *Journal of Cognitive Neuroscience, 9*, 254–265. https://doi.org/10.1162/jocn.1997.9.2.254

Chew, S. L. (2021). An advance organizer for student learning: Choke points and pitfalls in studying. *Canadian Psychology* [Online Publication Ahead of Print]. 10.1037/cap0000290.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Davis, M., & Hult, R. E. (1997). Effects of writing summaries as a generative learning activity during note taking. *Teaching of Psychology, 24*(1), 47–49. https://doi.org/10.1177/009862839702400112

*Desselle, S. P., & Shane, P.A. (2018). Laptop versus longhand note taking in a professional doctoral course: Student performance, attitudes, and behaviors. *INNOVATIONS in Pharmacy, 9*(3), 1-12. 10.24926/iip.v9i3.1392.

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

Dontre, A. J. (2020). The influence of technology on academic distraction: A review. *Human Behavior and Emerging Technologies, 1–12*. https://doi.org/10.1002/hbe2.229

Dumitru, D. (2019). Creating meaning. The importance of Arts, Humanities and Culture for critical thinking development. *Studies in Higher Education, 44*(5), 870–879. https://doi.org/10.1080/03075079.2019.1586345

Duran, K. S., & Frederick, C. M. (2013). Information comprehension: Handwritten vs. typed notes. Retrieved from *Undergraduate Research Journal for the Human Sciences, 12*(1) https://www.kon.org/urc/urc_research_journal12.html.

*Eason, T. (2017). *Various methods of note-taking and how they compare in terms of information retention* [Unpublished master's thesis]. East Carolina University.

Edmondson, K. M., & Novak, J. D. (1993). The interplay of scientific epistemological views, learning strategies, and attitudes of college students. *Journal of Research in Science Teaching, 30*(6), 547–559. https://doi.org/10.1002/(ISSN)1098-273610.1002/tea.v30:610.1002/tea.3660300604

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

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

Fulton, S., Schweitzer, D., Scharff, L., & Boleng, J. (2011). Demonstrating the impact of multitasking in the classroom. *Frontiers in Education Conference, FIE.*. https://doi.org/10.1109/FIE.2011.6142991

Goodwin, B. (2018). The magic of writing stuff down. *Educational Leadership, April 2018,* 78-79.

Gurley, D. L. (2018). *The relationship between note-taking method and grade point average when controlling for ACT score and self-regulation.* Unpublished doctoral dissertation, Department of Higher Education, University of Mississippi.
Hedges, L. V., & Becker, B. J. (1986). Statistical methods in the meta-analysis of research on gender differences. In J. Hyde, & M. C. Linn (Eds.), *The psychology of gender: Advance through meta-analysis* (pp. 14–50). John Hopkins University Press.
Hembrooke, H., & Gay, G. (2003). The laptop and the lecture: The effects of multitasking in learning environments. *Journal of Computing in Higher Education, 15*(1), 46–64. https://doi.org/10.1007/BF02940852
Horwitz, S. M. (2017). Is note-taking more effective with a keyboard or a pen?. *Unpublished Master's Thesis*. University of Colorado Boulder.
Igo, L. B., Riccomini, P. J., Bruning, R. H., & Pope, G. G. (2006). How should middle-school students with LD approach online note taking? A mixed-methods study. *Learning Disability Quarterly, 29*, 89–100. https://doi.org/10.2307/30035537
*Jackson, E. (2016). *Comparing analog and digital note-taking among middle school students* [Unpublished Doctoral dissertation]. Northcentral University.
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*, Article 106264. https://doi.org/10.1016/j.chb.2020.106264
Jansen, R. S., Lakens, D., & IJsselsteijn, W. A. (2017). An integrative review of the cognitive costs and benefits of note-taking. *Educational Research Review, 22,* 223-233. 10.1016/j.edurev.2017.10.001.
Jiang, Y., Clarke-Midura, J., Keller, B., Baker, R. S., Paquette, L., & Ocumpaugh, J. (2018). Note-taking and science inquiry in an open-ended learning environment. *Contemporary Educational Psychology, 55*, 12–29. https://doi.org/10.1016/j.cedpych.2018.08.004
Johnson, B. T., & Hennessy, E. A. (2019). Systematic reviews and meta-analyses in the health sciences: Best practice methods for research syntheses. *Social Science & Medicine, 233*, 237–251. https://doi.org/10.1016/j.socscimed.2019.05.035
Kane, M. J., Carruth, N. P., Lurquin, J. H., Silvia, P. J., Smeekens, B. A., von Bastian, C. C., & Miyake, A. (2021). Individual differences in task-unrelated thought in university classrooms. *Memory & Cognition, 49*(6), 1247–1266. https://doi.org/10.3758/s13421-021-01156-3
Kay, R., Benzimra, D., & Li, J. (2017). Exploring factors that influence technology-based distractions in bring your own device classrooms. *Journal of Educational Computing Research*, *55*(7), 974-995. 10.1177%2F0735633117690004.
*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), 2080–2084. 10.5688/ajpe7535.
Kiewra, Kenneth A. (1985). Investigating notetaking and review: A depth of processing alternative. *Educational Psychologist, 20*(1), 23–32. https://doi.org/10.1207/s15326985ep2001_4
Kiewra, K., Dubois, N., Christian, D., McShane, A., Meyerhoffer, M., & Roskelley, D. (1991). Note-taking functions and techniques. *Journal of Educational Psychology, 83*, 240–245. https://doi.org/10.1037/0022-0663.83.2.240
*Kirkland, K. M. (2016). *The effect of note taking media and preference on the cognitive processes involved in learning.* [Undergraduate Honors Theses], University of Colorado Boulder.
*Kodaira, Y. (2017). *An investigation of cognitive processes associated with notetaking and notes-review* [Unpublished doctoral dissertation]. Columbia University.
*Kutta, T. J. (2017). *Note-taking and technology [Unpublished doctoral dissertation]*. Florida State University.
Lin, L., & Bigenho, C. (2011). Note-Taking and memory in different media environments. *Computers in the Schools, 28*(3), 200–216. https://doi.org/10.1080/07380569.2011.594989
Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. *Sage Publications*.
Lockhart, R. S., & Craik, F. I. (1990). Levels of processing: A retrospective commentary on a framework for memory research. *Canadian Journal of Psychology, 44*, 87–112. https://doi.org/10.1037/h0084237
Luo, L., Kiewra, K. A., Flanigan, A. E., & Peteranetz, M. S. (2018). Laptop versus longhand note taking: Effects on lecture notes and achievement. *Instructional Science, 46*, 947–971. https://doi.org/10.1007/s11251-018-9458-0
McCann, Lee I., Ebert, Alexandria R., Oechsner, Markus, Immel, Kathy R., & Kadah-Ammeter, Tammy L. (2019). Studying for introductory psychology: Differences across technical college, 2-, and 4-year institutions. *Teaching of Psychology, 46*(4), 312–318. https://doi.org/10.1177/0098628319872591
*Mitchell, A., & Zheng, L. (2017). *Examining longhand vs. laptop debate: Evidence from a replication.* [Unpublished report]. Retrieved from https://aisel.aisnet.org/amcis2017/Replication/Presentations/2/.
Morehead, K., Dunlosky, J., Rawson, K. A., Blasiman, R., & Hollis, R. B. (2019). Note-taking habits of 21st century college students: Implications for student learning, memory, and achievement. *Memory, 27*, 807–819. https://doi.org/10.1080/09658211.2019.1569694
Mueller, P. A., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. *Psychological Science, 25*, 1159–1168. https://doi.org/10.1177/0956797614524581
Mueller, P. A., & Oppenheimer, D. M. (2018). "The pen is mightier than the keyboard: Advantages of longhand over laptop note taking": Corrigendum. *Psychological Science, 29*(9), 1565–1568. https://doi.org/10.1177/0956797618781773
Olive, Thierry, & Barbier, Marie-Laure (2017). Processing time and cognitive effort of longhand note-taking when reading and summarizing a structured or linear text. *Written Communication, 34*(2), 224–246. https://doi.org/10.1177/0741088317699898
Pabuccu, Aybuke, & Erduran, Sibel (2017). Beyond rote learning in organic chemistry: The infusion and impact of argumentation in tertiary education. *International Journal of Science Education, 39*(9), 1154–1172. https://doi.org/10.1080/09500693.2017.1319988
Patterson, R. W., & Patterson, R. M. (2017). Computers and productivity: Evidence from laptop use in the college classroom. *Economics of Education Review, 57*, 66–79. https://doi.org/10.1016/j.econedurev.2017.02.004
**Pettit-O'Malley, K. L., Liesz, T. J., & Sisodiya, S. R. (2017). The relative efficacy of handwritten versus electronic student classroom notes. *Business Education Innovation Journal, 9*(2), 110–197.
Peverly, S. T., Ramaswamy, V., Brown, C., Sumowski, J., Alidoost, M., & Garner, J. (2007). What predicts skill in lecture note taking? *Journal of Educational Psychology, 99*, 167–180. https://doi.org/10.1037/0022-0663.99.1.167
Piolat, Annie, Olive, Thierry, & Kellogg, Ronald T. (2005). Cognitive effort during note taking. *Applied Cognitive Psychology, 19*(3), 291–312. https://doi.org/10.1002/(ISSN)1099-072010.1002/acp.v19:310.1002/acp.1086
*Quade, A. M. (1995). *A comparison of on-line and traditional paper and pencil notetaking methods during computer-delivered instruction.* [Unpublished report]. Retrieved from https://eric.ed.gov/?id=ED383330.
Rana, N. P., Slade, E., Kitching, S., & Dwivedi, Y. K. (2019). The IT way of loafing in class: Extending the theory of planned behavior (TPB) to understand students' cyberslacking intentions. *Computers in Human Behavior, 101*, 114–123. https://doi.org/10.1016/j.chb.2019.07.022
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.
Rawson, K. A., Vaughn, K. E., Walsh, M., & Dunlosky, J. (2018). Investigating and explaining the effects of successive relearning on long-term retention. *Journal of Experimental Psychology: Applied, 24*(1), 57–71. https://doi.org/10.1037/xap0000146
*Richards, T., Peverly, S., Wolf, A., Abbott, R., Tanimoto, S., Thompson, R., Nagy, W., & Berninger, V. (2016). Idea units in notes and summaries for read texts by keyboard and pencil in middle childhood students with specific learning disabilities: Cognitive and brain findings. *Trends in Neuroscience and Education, 5*(3), 146–155. https://doi.org/10.1016/j.tine.2016.07.005
Rodgers, Melissa A., & Pustejovsky, James E. (2020). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods. Advance online publication., 26*(2), 141–160.
Ritter, N. L. (2015). The need for meta-analytic thinking in educational technology research. (Doctoral dissertation). Retrieved from https://www.researchgate.net/profile/Nicola-Ritter-2/publication/291827867_The_Need_for_Meta-Analytic_Thinking_in_Educational_Technology/links/597a23ebaca272e8cc0d399c/The-Need-for-Meta-Analytic-Thinking-in-Educational-Technology.pdf.
*Schoen, I. (2012). Effects of method and context of note-taking on memory: Handwriting versus typing in lecture and textbook-reading contexts [Unpublished senior's thesis]. Pitzer College.
*Smoker, T. J., Murphy, C. E., & Rockwell, A. K. (2009). Comparing memory for handwriting versus typing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 53*(22), 1744–1747. 10.1177/154193120905302218.
Stefanou, Candice, Hoffman, Lynn, & Vielee, Nicolette (2008). Note taking in the college classroom as evidence of generative learning. *Learning Environments Research, 11*(1), 1–17. https://doi.org/10.1007/s10984-007-9033-0
Sun, D., & Li, Y. (2019). Effectiveness of digital note-taking on students' performance in declarative, procedural and conditional knowledge learning. *International Journal of Emerging Technologies in Learning (iJET), 14*(18), 108–119. https://doi.org/10.3991/ijet.v14i18.10825
Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon.
Tesch, F., Coelho, D., & Drozdenko, R. (2011). We have met the enemy and he is us: Relative potencies of classroom distractions. *Business Education Innovation Journal, 3* (2), 13–19.
Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher, 31*(3), 25–32. https://doi.org/10.3102/0013189X031003025
Tipton, E., & Pustejovsky, J. (2015). Small-sample adjustments to multivariate hypothesis tests in robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics, 40*, 604–634. https://doi.org/10.3102/1076998615606099
Tulving, Endel, & Thomson, Donald M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352–373. https://doi.org/10.1037/h0020071
Unsworth, N., & McMillan, B. D. (2017). Attentional disengagements in educational contexts: A diary investigation of everyday mind-wandering and distraction. *Cognitive Research: Principles and Implications, 2*(1), 32. https://doi.org/10.1186/s41235-017-0070-7
*Urry, H. L. et al. (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,* 326-339. 10.31234/osf.io/vqyw6.
Van den Noortgate, Wim, López-López, José Antonio, Marín-Martínez, Fulgencio, & Sánchez-Meca, Julio (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods, 45*(2), 576–594. https://doi.org/10.3758/s13428-012-0261-6
Van den Noortgate, Wim, López-López, José Antonio, Marín-Martínez, Fulgencio, & Sánchez-Meca, Julio (2015). Meta-analysis of multiple outcomes: A multilevel approach. *Behavior Research Methods, 47*(4), 1274–1294. https://doi.org/10.3758/s13428-014-0527-2
Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. URL *Journal of Statistical Software, 36*, 1–48 http://www.jstatsoft.org/v36/i03/.

Waite, Bradley M., Lindberg, Rachel, Ernst, Brittany, Bowman, Laura L., & Levine, Laura E. (2018). Off-task multitasking, note-taking and lower- and higher-order classroom learning. *Computers and Education, 120*, 98–111. https://doi.org/10.1016/j.compedu.2018.01.007

Wammes, J. D., Boucher, P. O., Seli, P., Cheyne, J. A., & Smilek, D. (2016). Mind wandering during lectures I: Changes in rates across an entire semester. *Scholarship of Teaching and Learning in Psychology, 2*(1), 13–32. https://doi.org/10.1037/stl0000053

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

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

Zeamer, C., & Fox Tree, J. E. (2013). The process of auditory distraction: Disrupted attention and impaired recall in a simulated lecture environment. *Journal of experimental psychology. Learning, memory, and cognition, 39*(5), 1463–1472. https://doi.org/10.1037/a0032190

*Zyrowski, M. L. (2014). *Digital annotation tools and comprehension* [Unpublished Doctoral dissertation]. Oakland University.