

Corrigendum: The Pen Is Mightier Than the Keyboard: Advantages of Longhand Over Laptop Note Taking

Psychological Science

1–4

© The Author(s) 2018

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0956797618781773

www.psychologicalscience.org/PS



Original article: 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. doi:10.1177/0956797614524581

This Corrigendum corrects five categories of errors in the original article:

- The z scores reported in the data files uploaded to the Open Science Framework (OSF) were calculated prior to having to remove 1 participant in Study 1, 2 participants in Study 2, and 9 participants in Study 3 for various reasons (which were reported in the original article). We had removed those participants from the data files on OSF but had failed to recalculate the z scores without those participants included.
- In the original article, the z scores from Study 1 were calculated using an index score across lectures (a perfect score would be 1 point per question; 10 points total), whereas the z scores in Studies 2 and 3 were calculated using the raw data (in which different questions had different point values). The article erroneously indicated that the index-score approach was used for all three studies. We will now report all results using index scores; the pattern of results is the same for both measures.
- Moreover, the data in Study 2 were z -scored within lectures, which limited the inferential power of the analyses. We will now report results from Study 2 using z scores across lectures. The patterns of results do not appreciably change.
- The article mistakenly reported the degrees of freedom for error for the interaction of condition and lecture in the first two studies (55 and 89 for Study 1 and Study 2, respectively) instead of the degrees of freedom for error for condition (4.01 and 4.09 for Study 1 and Study 2, respectively, due to the use of mixed and random-effects analyses).
- Some of the effect sizes were reported as η_p^2 values when they were actually η^2 values. The effect sizes reported in this Corrigendum (and

which will be corrected in the article) are the η_p^2 values, as presented in SPSS.

New data files with the corrected z scores, as well as SPSS syntax for the corrected analyses, have been uploaded to OSF (<https://osf.io/crsiz>). A file with the syntax for the originally uploaded files with the z -scoring errors has also been included (see the associated Wiki for detailed descriptions). Additionally, because of the large number of experimenter degrees of freedom in these analyses (e.g., whether we used index or raw scores), to demonstrate that the effects are robust beyond specific analysis decisions, we have posted a series of analyses showing how the results change across different analysis strategies. We regret the errors and inconsistencies in the original manuscript but are heartened that regardless of how the data are analyzed, the results remain qualitatively the same.

Results that are now being corrected in the article are as follows.

Study 1

On page 1161, in the Results section of Study 1, the values in the second half of the first paragraph will be changed as follows (the text and overall conclusions remain the same):

On factual-recall questions, participants performed equally well across conditions (laptop: $M = -0.006$, $SD = 1.00$; longhand: $M = 0.05$, $SD = 1.01$), $F(1, 4.01) = 0.046$, $p = .841$. However, on conceptual-application questions, laptop participants performed significantly worse ($M = -0.178$, $SD = 0.900$) than longhand participants ($M = 0.162$, $SD = 1.07$), $F(1, 4.09) = 8.05$, $p = .046$, $\eta_p^2 = .66$ (see Fig. 1).⁵ Which lecture participants saw also affected performance on conceptual-application

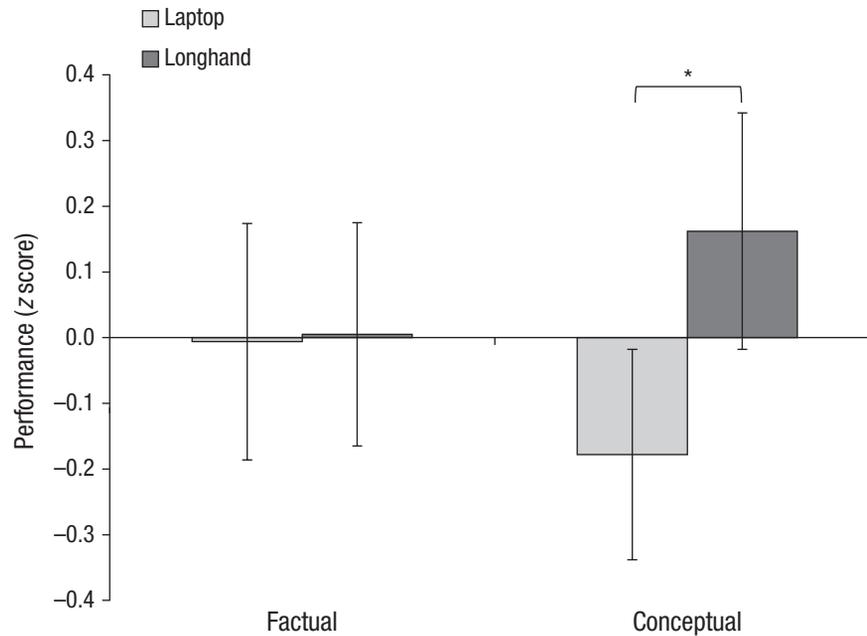


Fig. 1. Mean performance on factual-recall and conceptual-application questions as a function of note-taking condition (Study 1). Performance was converted to an index score where each question was worth 1 point, and the totals were z-scored across lectures. The asterisk indicates a significant difference between conditions ($p < .05$). Error bars indicate standard errors of the mean.

questions, $F(4, 4) = 7.11, p = .042, \eta_p^2 = .88$; however, there was no significant interaction between lecture and note-taking medium, $F(4, 55) = 0.259, p = .90$.

Figure 1 will also be corrected to show the new means, and the caption will be changed as indicated.

In the original article, we did not report the null effect of performance across lectures on factual questions. To fill that omission, we now report the following at the end of the first paragraph on p. 1161: “There was no significant difference in performance on factual questions across lectures, $F(4, 4) = 1.57, p = .33$.”

Study 2

On page 1162, the first paragraph of the “Laptop versus longhand performance” subsection will be replaced with the following (the overall conclusions remain the same):

Responses were scored by raters blind to condition. Replicating our original finding, results showed that on conceptual-application questions, longhand participants performed better (z -score $M = 0.24, SD = 1.11$) than laptop-nonintervention participants (z -score $M = -0.17, SD = 0.88$), $F(1, 6.60) = 19.65, p = .003, \eta_p^2 = .75$. Scores for laptop-intervention participants (z -score $M = -0.11, SD = 1.02$) did not significantly differ from those for

either laptop-nonintervention ($p = .91$) or longhand ($p = .29$) participants. The pattern of data for factual questions was similar, though the differences were not significant (longhand: z -score $M = 0.025, SD = 0.97$; laptop intervention: z -score $M = 0.063, SD = 1.05$; laptop nonintervention: z -score $M = -0.089, SD = 0.99$), $F(1, 4.54) = 4.08, p = .11$ (see Fig. 4).⁸ There was a significant difference in conceptual performance across lectures, $F(4, 4) = 19.87, p = .007, \eta_p^2 = .95$, but the interaction was not significant, $F(4, 89) = 0.138, p = .97$. There was a significant difference in factual performance across lectures, $F(4, 4) = 14.59, p = .012, \eta_p^2 = .94$, but the interaction was not significant, $F(4, 89) = 0.439, p = .66$.

Figure 4 will also be corrected to show the new means, and the caption will be changed as indicated.

In addition, a new endnote (Note 9) will be added at the end of the Results section for Study 2 (p. 1163). The text of this note will read as follows:

Estimating effect sizes in mixed models is problematic (for more details, see, e.g., Judd, Westfall, & Kenny, 2017). Thus, while the η_p^2 s reported above for Studies 1 and 2 (ranging from .66 to .95) are what SPSS reports, more informative estimates of effect sizes would be much smaller, ranging from .03 to .11.

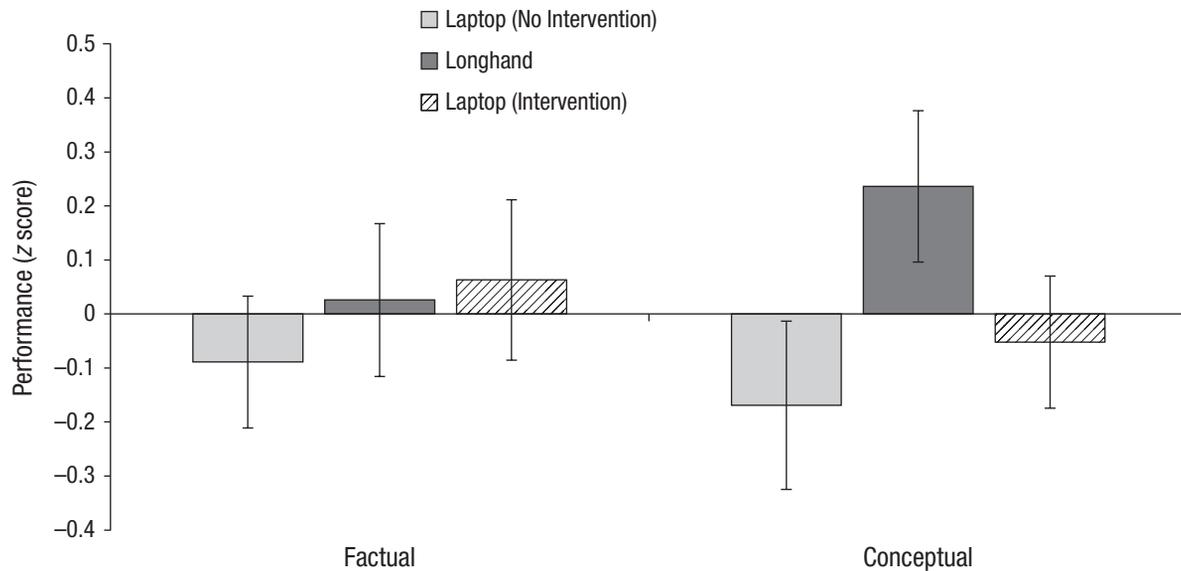


Fig. 4. Mean performance on factual-recall and conceptual-application questions as a function of note-taking condition (Study 2). Performance was converted to an index score where each question was worth 1 point, and the totals were z-scored across lectures. Error bars indicate standard errors of the mean.

Study 3

Results for Study 3 were calculated using z scores of raw rather than index scores. The overall conclusions stand, but the two paragraphs in the “Laptop versus longhand performance” subsection (p. 1164) will be changed as follows:

Across all question types, there were no main effects of note-taking medium or opportunity to study. However, there was a significant interaction between these two variables, $F(1, 105) = 5.62, p = .02, \eta_p^2 = .05$. Participants who took longhand notes and were able to study them performed significantly better (z -score $M = 0.45$) than participants in any of the other conditions (z -score $M_s = -0.25, -0.02, -0.20$), $t(105) = 2.86, p = .005, d = 0.56$ (see Fig. 5).

Collapsing questions about facts and seductive details into a general measure of “factual” performance, we found a significant main effect of note-taking medium, $F(1, 105) = 4.05, p = .047, \eta_p^2 = .037$, and of opportunity to study, $F(1, 105) = 11.49, p = .001, \eta_p^2 = .09$, but this was qualified by a significant interaction, $F(1, 105) = 5.45, p = .021, \eta_p^2 = .05$. Again, participants in the longhand-study condition (z -score $M = 0.68$) outperformed the other participants (z -score $M_s = -0.09, -0.28, -0.34$), $t(105) = 4.50, p < .001, d = 0.88$. Among students who had the opportunity to study,

longhand note takers did significantly better than laptop note takers, $t(53) = 2.77, p = .008$.

Collapsing performance on conceptual, inferential, and application questions into a general “conceptual” measure revealed no significant main effects, but again there was a significant interaction between note-taking medium and studying. There was a significant interaction for conceptual questions, $F(1, 105) = 4.35, p = .04, \eta_p^2 = .04$. Among students who had the opportunity to study, longhand note takers did significantly better than laptop note takers, $t(53) = 2.32, p = .024, d = 0.64$ (for raw means, see Table 2).

The following sentence will be added for clarification to the note to Table 2: “Although raw scores are given here, z scores of the index scores were used in the analysis.” Figure 5 will also be corrected to show the new means, and the caption will be changed as indicated.

A new endnote (Note 10) will be added at the end of the Results section for Study 3 (p. 1166). The text of this note will read as follows:

Participants who took laptop or longhand notes but who did not have the opportunity to study did not score significantly differently on either factual, $t(52) = 0.26, p = .795, d = 0.07$, or conceptual, $t(52) = 0.77, p = .442, d = 0.21$, questions.

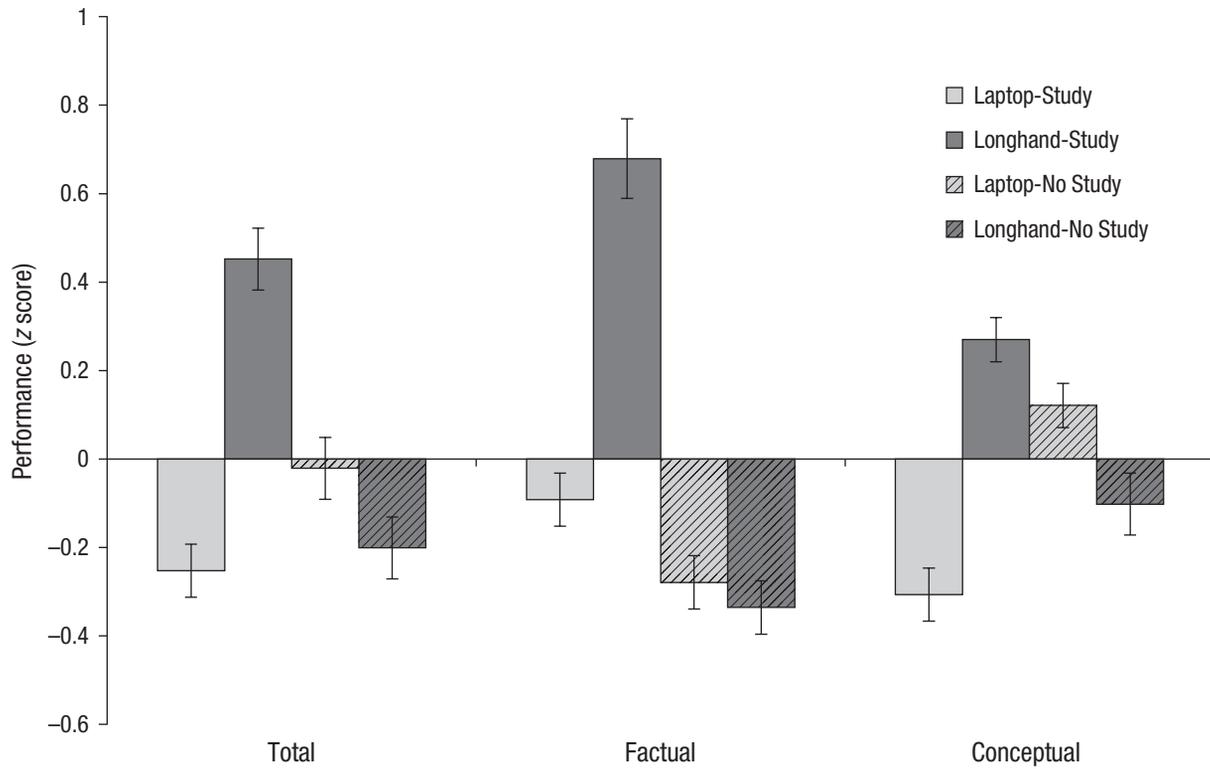


Fig. 5. Mean performance on factual-recall and conceptual-application questions as a function of note-taking condition and opportunity to study (Study 3). Performance was converted to an index score where each question was worth 1 point, and the totals were z -scored across lectures. Error bars indicate standard errors of the mean.

References

The following reference will be added to the References section:

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, 68, 601–625.

Corresponding Author

The Corresponding Author will be changed to Daniel M. Oppenheimer (p. 1159). The new contact information is as follows: Carnegie Mellon University, Department of Social and Decision Sciences and Department of Psychology, Porter Hall 208, 5000 Forbes Ave., Pittsburgh, PA 15213. E-mail: oppenheimer@cmu.edu

The Pen Is Mightier Than the Keyboard: Advantages of Longhand Over Laptop Note Taking



Pam A. Mueller¹ and Daniel M. Oppenheimer²

¹Princeton University and ²University of California, Los Angeles

Psychological Science
 2014, Vol. 25(6) 1159–1168
 © The Author(s) 2014
 Reprints and permissions:
sagepub.com/journalsPermissions.nav
 DOI: 10.1177/0956797614524581
pss.sagepub.com


Abstract

Taking notes on laptops rather than in longhand is increasingly common. Many researchers have suggested that laptop note taking is less effective than longhand note taking for learning. Prior studies have primarily focused on students' capacity for multitasking and distraction when using laptops. The present research suggests that even when laptops are used solely to take notes, they may still be impairing learning because their use results in shallower processing. In three studies, we found that students who took notes on laptops performed worse on conceptual questions than students who took notes longhand. We show that whereas taking more notes can be beneficial, laptop note takers' tendency to transcribe lectures verbatim rather than processing information and reframing it in their own words is detrimental to learning.

Keywords

academic achievement, cognitive processes, memory, educational psychology, open data, open materials

Received 5/11/13; Revision accepted 1/16/14

The use of laptops in classrooms is controversial. Many professors believe that computers (and the Internet) serve as distractions, detracting from class discussion and student learning (e.g., Yamamoto, 2007). Conversely, students often self-report a belief that laptops in class are beneficial (e.g., Barak, Lipson, & Lerman, 2006; Mitra & Steffensmeier, 2000; Skolnick & Puzo, 2008). Even when students admit that laptops are a distraction, they believe the benefits outweigh the costs (Kay & Lauricella, 2011). Empirical research tends to support the professors' view, finding that students using laptops are not on task during lectures (Kay & Lauricella, 2011; Kraushaar & Novak, 2010; Skolnick & Puzo, 2008; Sovern, 2013), show decreased academic performance (Fried, 2008; Grace-Martin & Gay, 2001; Kraushaar & Novak, 2010), and are actually less satisfied with their education than their peers who do not use laptops in class (Wurst, Smarkola, & Gaffney, 2008).

These correlational studies have focused on the capacity of laptops to distract and to invite multitasking. Experimental tests of immediate retention of class material have also found that Internet browsing impairs performance (Hembrooke & Gay, 2003). These findings are important but relatively unsurprising, given the literature

on decrements in performance when multitasking or task switching (e.g., Iqbal & Horvitz, 2007; Rubinstein, Meyer, & Evans, 2001).

However, even when distractions are controlled for, laptop use might impair performance by affecting the manner and quality of in-class note taking. There is a substantial literature on the general effectiveness of note taking in educational settings, but it mostly predates laptop use in classrooms. Prior research has focused on two ways in which note taking can affect learning: encoding and external storage (see DiVesta & Gray, 1972; Kiewra, 1989). The encoding hypothesis suggests that the processing that occurs during the act of note taking improves learning and retention. The external-storage hypothesis touts the benefits of the ability to review material (even from notes taken by someone else). These two theories are not incompatible; students who both take and review

Corresponding Author:

Daniel M. Oppenheimer, Carnegie Mellon University, Department of Social and Decision Sciences and Department of Psychology, Porter Hall 208, 5000 Forbes Ave., Pittsburgh, PA 15213
 E-mail: oppenheimer@cmu.edu

their notes (as most do) likely profit from both approaches (Kiewra, 1985).

The beneficial external-storage effect of notes is robust and uncontroversial (Kiewra, 1989). The encoding hypothesis has been supported by studies finding positive effects of note taking in the absence of review (e.g., Aiken, Thomas, & Shennum, 1975; Bretzing & Kulhavy, 1981; Einstein, Morris, & Smith, 1985); however, other results have been more mixed (see Kiewra, 1985; Kobayashi, 2005, for reviews). This inconsistency may be a result of moderating factors (Kobayashi, 2005), potentially including one's note-taking strategy.

Note taking can be generative (e.g., summarizing, paraphrasing, concept mapping) or nongenerative (i.e., verbatim copying). Verbatim note taking has generally been seen to indicate relatively shallow cognitive processing (Craik & Lockhart, 1972; Kiewra, 1985; Van Meter, Yokoi, & Pressley, 1994). The more deeply information is processed during note taking, the greater the encoding benefits (DiVesta & Gray, 1973; Kiewra, 1985). Studies have shown both correlationally (Aiken et al., 1975; Slotte & Lonka, 1999) and experimentally (Bretzing & Kulhavy, 1979; Igo, Bruning, & McCrudden, 2005) that verbatim note taking predicts poorer performance than nonverbatim note taking, especially on integrative and conceptual items.

Laptop use facilitates verbatim transcription of lecture content because most students can type significantly faster than they can write (Brown, 1988). Thus, typing may impair the encoding benefits seen in past note-taking studies. However, the ability to transcribe might improve external-storage benefits.

There has been little research directly addressing potential differences in laptop versus longhand note taking, and the existing studies do not allow for natural variation in the amount of verbatim overlap (i.e., the amount of text in common between a lecture and students' notes on that lecture). For example, Bui, Myerson, and Hale (2013) found an advantage for laptop over longhand note taking. However, their results were driven by a condition in which they explicitly instructed participants to transcribe content, rather than allowing them to take notes as they would in class. Lin and Bigenho (2011) used word lists as stimuli, which also ensured that all note taking would be verbatim. Therefore, these studies do not speak to real-world settings, where laptop and longhand note taking might naturally elicit different strategies regarding the extent of verbatim transcription.¹ Moreover, these studies only tested immediate recall, and exclusively measured factual (rather than conceptual) knowledge, which limits generalizability (see also Bohay, Blakely, Tamplin, & Radvansky, 2011; Quade, 1996). Previous studies have shown that detriments due to verbatim note taking are more prominent for

conceptual than for factual items (e.g., Bretzing & Kulhavy, 1979).

Thus, we conducted three experiments to investigate whether taking notes on a laptop versus writing longhand affects academic performance, and to explore the potential mechanism of verbatim overlap as a proxy for depth of processing.

Study 1

Participants

Participants were 67 students (33 male, 33 female, 1 unknown) from the Princeton University subject pool. Two participants were excluded, 1 because he had seen the lecture serving as the stimulus prior to participation, and 1 because of a data-recording error.

Materials

We selected five TED Talks (<https://www.ted.com/talks>) for length (slightly over 15 min) and to cover topics that would be interesting but not common knowledge.² Laptops had full-size (11-in. × 4-in.) keyboards and were disconnected from the Internet.

Procedure

Students generally participated 2 at a time, though some completed the study alone. The room was preset with either laptops or notebooks, according to condition. Lectures were projected onto a screen at the front of the room. Participants were instructed to use their normal classroom note-taking strategy, because experimenters were interested in how information was actually recorded in class lectures. The experimenter left the room while the lecture played.

Next, participants were taken to a lab; they completed two 5-min distractor tasks and engaged in a taxing working memory task (viz., a reading span task; Unsworth, Heitz, Schrock, & Engle, 2005). At this point, approximately 30 min had elapsed since the end of the lecture. Finally, participants responded to both factual-recall questions (e.g., "Approximately how many years ago did the Indus civilization exist?") and conceptual-application questions (e.g., "How do Japan and Sweden differ in their approaches to equality within their societies?") about the lecture and completed demographic measures.³

The first author scored all responses blind to condition. An independent rater, blind to the purpose of the study and condition, also scored all open-ended questions. Initial interrater reliability was good ($\alpha = .89$); score disputes between raters were resolved by discussion. Longhand notes were transcribed into text files.

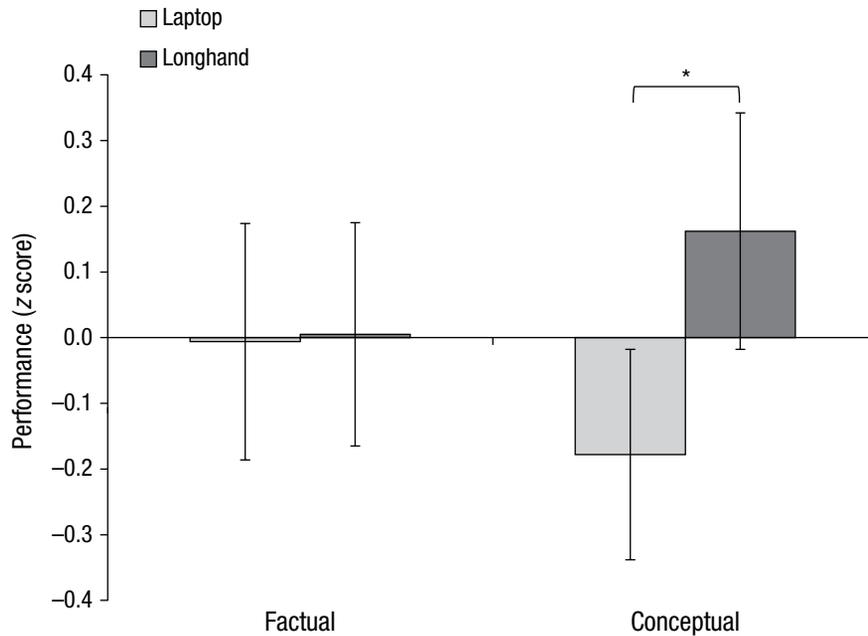


Fig. 1. Mean performance on factual-recall and conceptual-application questions as a function of note-taking condition (Study 1). Performance was converted to an index score where each question was worth 1 point, and the totals were z-scored across lectures. The asterisk indicates a significant difference between conditions ($p < .05$). Error bars indicate standard errors of the mean.

Results and discussion

Laptop versus longhand performance. Mixed fixed- and random-effects analyses of variance were used to test differences, with note-taking medium (laptop vs. longhand) as a fixed effect and lecture (which talk was viewed) as a random effect. We converted the raw data to z scores because the lecture assessments varied in difficulty and number of points available; however, results did not differ when raw scores were analyzed.⁴ On factual-recall questions, participants performed equally well across conditions (laptop: $M = -0.006$, $SD = 1.00$; longhand: $M = 0.05$, $SD = 1.01$), $F(1, 4.01) = 0.046$, $p = .841$. However, on conceptual-application questions, laptop participants performed significantly worse ($M = -0.178$, $SD = 0.900$) than longhand participants ($M = 0.162$, $SD = 1.07$), $F(1, 4.09) = 8.05$, $p = .046$, $\eta_p^2 = .66$ (see Fig. 1).⁵ Which lecture participants saw also affected performance on conceptual-application questions, $F(4, 4) = 7.11$, $p = .042$, $\eta_p^2 = .88$; however, there was no significant interaction between lecture and note-taking medium, $F(4, 55) = 0.259$, $p = .90$. There was no significant difference in performance on factual questions across lectures, $F(4, 4) = 1.57$, $p = .33$.

Content analysis. There were several qualitative differences between laptop and longhand notes.⁶ Participants who took longhand notes wrote significantly fewer words ($M = 173.4$, $SD = 70.7$) than those who typed ($M = 309.6$, $SD = 116.5$), $t(48.58) = -5.63$, $p < .001$,

$d = 1.4$, corrected for unequal variances (see Fig. 2). A simple n -gram program measured the extent of textual overlap between student notes and lecture transcripts. It compared each one-, two-, and three-word chunk of text in the notes taken with each one-, two-, and three-word chunk of text in the lecture transcript, and reported a percentage of matches for each. Using three-word chunks (3-grams) as the measure, we found that laptop

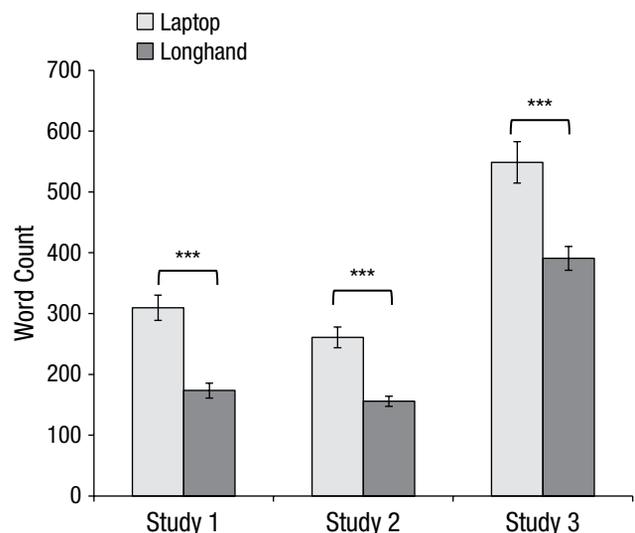


Fig. 2. Number of words written by students using laptops and notebooks in Studies 1, 2, and 3. Asterisks indicate a significant difference between conditions ($p < .001$). Error bars indicate standard errors of the mean.

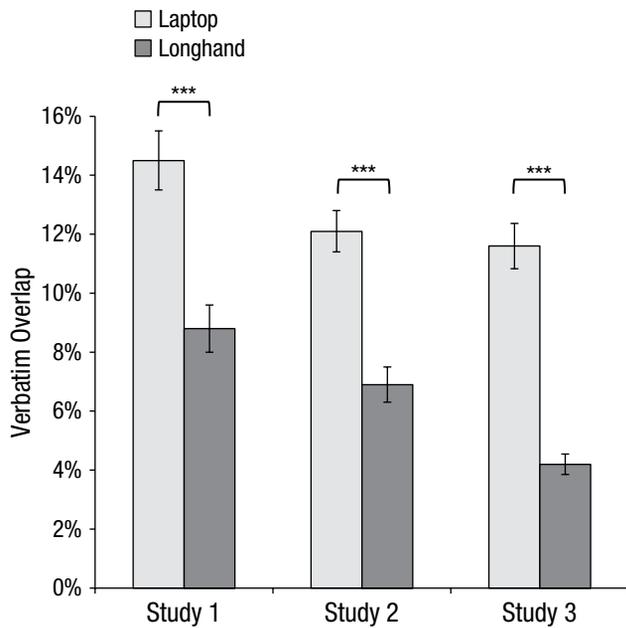


Fig. 3. Percentage of verbatim overlap between student notes and lecture transcripts in Studies 1, 2, and 3 as a function of note-taking condition. Verbatim overlap was measured using 3-grams (i.e., by comparing three-word chunks of text in the student notes and lecture transcripts). Error bars indicate standard errors of the mean.

notes contained an average of 14.6% verbatim overlap with the lecture ($SD = 7.3\%$), whereas longhand notes averaged only 8.8% ($SD = 4.8\%$), $t(63) = -3.77$, $p < .001$, $d = 0.94$ (see Fig. 3); 2-grams and 1-grams also showed significant differences in the same direction.

Overall, participants who took more notes performed better, $\beta = 0.34$, $p = .023$, partial $R^2 = .08$. However, those whose notes had less verbatim overlap with the lecture also performed better, $\beta = -0.43$, $p = .005$, partial $R^2 = .12$. We tested a model using word count and verbatim overlap as mediators of the relationship between note-taking medium and performance using Preacher and Hayes's (2004) bootstrapping procedure. The indirect effect is significant if its 95% confidence intervals do not include zero. The full model with note-taking medium as the independent variable and both word count and verbatim overlap as mediators was a significant predictor of performance, $F(3, 61) = 4.25$, $p = .009$, $R^2 = .17$. In the full model, the direct effect of note-taking medium remained a marginally significant predictor, $b = 0.54$ ($\beta = 0.27$), $p = .07$, partial $R^2 = .05$; both indirect effects were significant. Longhand note taking negatively predicted word count, and word count positively predicted performance, indirect effect = -0.57 , 95% confidence interval (CI) = $[-1.03, -0.20]$. Longhand note taking also negatively predicted verbatim overlap, and verbatim overlap negatively predicted performance, indirect effect = 0.34 , 95% CI = $[0.14, 0.71]$. Normal theory tests provided identical conclusions.⁷

This study provides initial experimental evidence that laptops may harm academic performance even when used as intended. Participants using laptops are more likely to take lengthier transcription-like notes with greater verbatim overlap with the lecture. Although taking more notes, thereby having more information, is beneficial, mindless transcription seems to offset the benefit of the increased content, at least when there is no opportunity for review.

Study 2

Because the detrimental effects of laptop note taking appear to be due to verbatim transcription, perhaps instructing students not to take verbatim notes could ameliorate the problem. Study 2 aimed to replicate the findings of Study 1 and to determine whether a simple instructional intervention could reduce the negative effects of laptop note taking. Moreover, we sought to show that the effects generalize to a different student sample.

Participants

Participants were students (final $N = 151$; 35 male) from the University of California, Los Angeles Anderson Behavioral Lab subject pool. Two participants were removed because of data-collection errors. Participants were paid \$10 for 1 hr of participation.

Procedure

Participants completed the study in groups. Each participant viewed one lecture on an individual monitor while wearing headphones. Stimuli were the same as in Study 1. Participants in the laptop-nonintervention and longhand conditions were given a laptop or pen and paper, respectively, and were instructed, "We're doing a study about how information is conveyed in the classroom. We'd like you to take notes on a lecture, just like you would in class. Please take whatever kind of notes you'd take in a class where you expected to be tested on the material later—don't change anything just because you're in a lab."

Participants in the laptop-intervention condition were instructed, "We're doing a study about how information is conveyed in the classroom. We'd like you to take notes on a lecture, just like you would in class. People who take class notes on laptops when they expect to be tested on the material later tend to transcribe what they're hearing without thinking about it much. Please try not to do this as you take notes today. Take notes in your own words and don't just write down word-for-word what the speaker is saying."

Participants then completed a typing test, the Need for Cognition scale (Cacioppo & Petty, 1982), academic

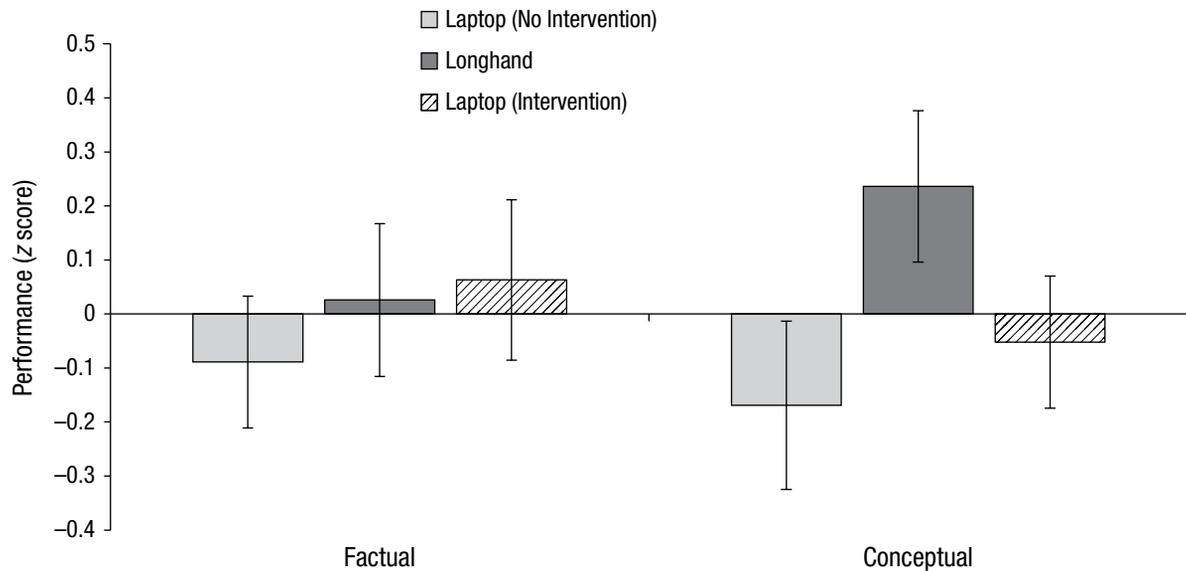


Fig. 4. Mean performance on factual-recall and conceptual-application questions as a function of note-taking condition (Study 2). Performance was converted to an index score where each question was worth 1 point, and the totals were z -scored across lectures. Error bars indicate standard errors of the mean.

self-efficacy scales, and a shortened version of the reading span task used in Study 1. Finally, they completed the same dependent measures and demographics as in Study 1. Longhand notes were transcribed, and all notes were analyzed with the n -grams program.

Results and discussion

Laptop versus longhand performance. Responses were scored by raters blind to condition. Replicating our original finding, results showed that on conceptual-application questions, longhand participants performed better (z -score $M = 0.24$, $SD = 1.11$) than laptop-nonintervention participants (z -score $M = -0.17$, $SD = 0.88$), $F(1, 6.60) = 19.65$, $p = .003$, $\eta_p^2 = .75$. Scores for laptop-intervention participants (z -score $M = -0.11$, $SD = 1.02$) did not significantly differ from those for either laptop-nonintervention ($p = .91$) or longhand ($p = .29$) participants. The pattern of data for factual questions was similar, though the differences were not significant (longhand: z -score $M = 0.025$, $SD = 0.97$; laptop intervention: z -score $M = 0.063$, $SD = 1.05$; laptop nonintervention: z -score $M = -0.089$, $SD = 0.99$), $F(1, 4.54) = 4.08$, $p = .11$ (see Fig. 4).⁸ There was a significant difference in conceptual performance across lectures, $F(4, 4) = 19.87$, $p = .007$, $\eta_p^2 = .95$, but the interaction was not significant, $F(4, 89) = 0.138$, $p = .97$. There was a significant difference in factual performance across lectures, $F(4, 4) = 14.59$, $p = .012$, $\eta_p^2 = .94$, but the interaction was not significant, $F(4, 89) = 0.439$, $p = .66$.

Participants' self-reported grade point average, SAT scores, academic self-efficacy, Need for Cognition scores, and reading span scores were correlated with performance on conceptual items, but were not significant covariates

when included in the overall analysis, so we will not discuss them further.

Content analysis. Participants who took longhand notes wrote significantly fewer words ($M = 155.9$, $SD = 59.6$) than those who took laptop notes without receiving an intervention ($M = 260.9$, $SD = 118.5$), $t(97) = -5.51$, $p < .001$, $d = 1.11$ (see Fig. 2), as well as less than those who took laptop notes after the verbal intervention ($M = 229.02$, $SD = 84.8$), $t(98) = -4.94$, $p < .001$, $d = 1.00$. Longhand participants also had significantly less verbatim overlap ($M = 6.9\%$, $SD = 4.2\%$) than laptop-nonintervention participants ($M = 12.11\%$, $SD = 5.0\%$), $t(97) = -5.58$, $p < .001$, $d = 1.12$ (see Fig. 3), or laptop-intervention participants ($M = 12.07\%$, $SD = 6.0\%$), $t(98) = -4.96$, $p < .001$, $d = 0.99$. The instruction to not take verbatim notes was completely ineffective at reducing verbatim content ($p = .97$).

Comparing longhand and laptop-nonintervention note taking, we found that for conceptual questions, participants taking more notes performed better, $\beta = 0.27$, $p = .02$, partial $R^2 = .05$, but those whose notes had less verbatim overlap also performed better, $\beta = -0.30$, $p = .01$, partial $R^2 = .06$, which replicates the findings of Study 1. We tested a model using word count and verbatim overlap as mediators of the relationship between note-taking medium and performance; it was a good fit, $F(3, 95) = 5.23$, $p = .002$, $R^2 = .14$. Again, both indirect effects were significant: Longhand note taking negatively predicted word count, and word count positively predicted performance, indirect effect = -0.34 , 95% CI = $[-0.56, -0.14]$. Longhand note taking also negatively predicted verbatim overlap, and verbatim overlap negatively predicted performance,

Table 1. Examples of Each Question Type Used in Study 3

Question type	Example
Factual	What is the purpose of adding calcium propionate to bread?
Seductive detail	What was the name of the cow whose cowpox was used to demonstrate the effectiveness of Edward Jenner's technique of inoculation against smallpox?
Conceptual	If a person's epiglottis was not working properly, what would be likely to happen?
Inferential	Sometimes bats die while they are sleeping. What will happen if a bat dies while it is hanging upside down?
Application	Psychologists have investigated a phenomenon known as "attitude inoculation," which works on the same principle as vaccination, and involves exposing people to weak arguments against a viewpoint they hold. What would this theory predict would happen if the person was later exposed to a strong argument against their viewpoint?

indirect effect = 0.19, 95% CI = [0.01, 0.49]. The direct effect of note-taking medium remained significant, $b = 0.58$ ($\beta = 0.30$), $p = .01$, partial $R^2 = .06$, so there is likely more at play than the two opposing mechanisms we identified here. When laptop (with intervention) was included as an intermediate condition, the pattern of effects remained the same, though the magnitude decreased; indirect effect of word count = -0.18 , 95% CI = $[-0.29, -0.08]$, indirect effect of verbatim overlap = 0.08, 95% CI = [0.01, 0.17].

The intervention did not improve memory performance above that for the laptop-nonintervention condition, but it was also not statistically distinguishable from memory in the longhand condition. However, the intervention was completely ineffective at reducing verbatim content, and the overall relationship between verbatim content and negative performance held. Thus, whereas the effect of the intervention on performance is ambiguous, any potential impact is unrelated to the mechanisms explored in this article.⁹

Study 3

Whereas laptop users may not be encoding as much information while taking notes as longhand writers are, they record significantly more content. It is possible that this increased external-storage capacity could boost performance on tests taken after an opportunity to study one's notes. Thus, in Study 3, we used a 2 (laptop, longhand) \times 2 (study, no study) design to investigate whether

the disadvantages of laptop note taking for encoding are potentially mitigated by enhanced external storage. We also continued to investigate whether there were consistent differences between responses to factual and conceptual questions, and additionally explored whether the note-taking medium affected transfer of learning of conceptual information to other domains (e.g., Barnett & Ceci, 2002).

Participants

Participants were students (final $N = 109$; 27 male) from the University of California, Los Angeles Anderson Behavioral Lab subject pool. One hundred forty-two participants completed Session 1 (presentation), but only 118 returned for Session 2 (testing). Of those 118, 8 participants were removed for not having taken notes or failing to respond to the test questions, and 1 was removed because of a recording error. Participant loss did not differ significantly across conditions. Participants were paid \$6 for the first session and \$7 for the second session.

Stimuli

Materials were adapted from Butler (2010). Four prose passages—on bats, bread, vaccines, and respiration—were read from a teleprompter by a graduate student acting as a professor at a lectern; two "seductive details" (i.e., "interesting, but unimportant, information"; Garner, Gillingham, & White, 1989, p. 41) were added to lectures that did not have them. Each filmed lecture lasted approximately 7 min.

Procedure

Participants completed the study in large groups. They were given either a laptop or pen and paper and were instructed to take notes on the lectures. They were told they would be returning the following week to be tested on the material. Each participant viewed all four lectures on individual monitors while wearing headphones.

When participants returned, those in the study condition were given 10 min to study their notes before being tested. Participants in the no-study condition immediately took the test. This dependent measure consisted of 40 questions, 10 on each lecture—two questions in each of five categories adapted from Butler (2010): facts, seductive details, concepts, same-domain inferences (inferences), and new-domain inferences (applications). See Table 1 for examples. Participants then answered demographic questions. All responses were scored by raters blind to condition. Longhand notes were transcribed, and all notes were analyzed using the n -grams program.

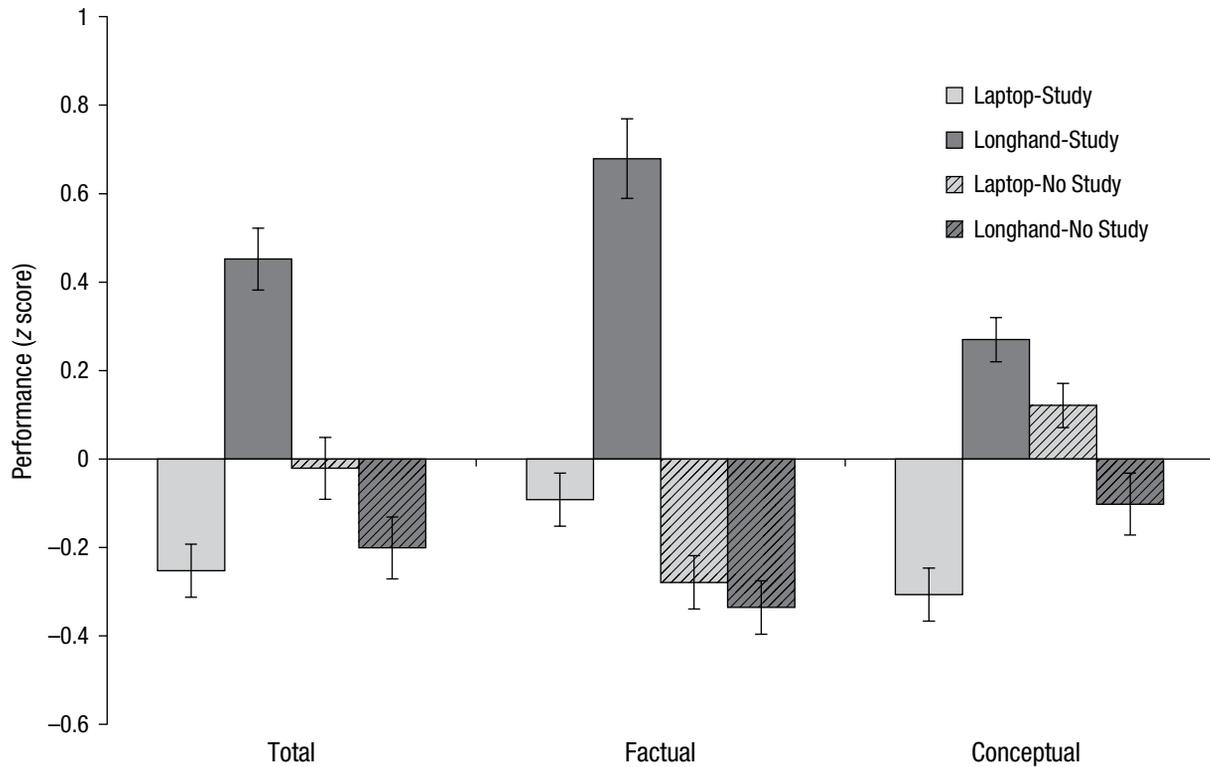


Fig. 5. Mean performance on factual-recall and conceptual-application questions as a function of note-taking condition and opportunity to study (Study 3). Performance was converted to an index score where each question was worth 1 point, and the totals were z-scored across lectures. Combined results for both question types are given separately. Error bars indicate standard errors of the mean.

Results

Laptop versus longhand performance. Across all question types, there were no main effects of note-taking medium or opportunity to study. However, there was a significant interaction between these two variables, $F(1, 105) = 5.62, p = .02, \eta_p^2 = .05$. Participants who took longhand notes and were able to study them performed significantly better (z-score $M = 0.45$) than participants in any of the other conditions (z-score $M_s = -0.25, -0.02, -0.20$), $t(105) = 2.86, p = .005, d = 0.56$ (see Fig. 5).

Collapsing questions about facts and seductive details into a general measure of “factual” performance, we found a significant main effect of note-taking medium, $F(1, 105) = 4.05, p = .047, \eta_p^2 = .037$, and of opportunity to study, $F(1, 105) = 11.49, p = .001, \eta_p^2 = .09$, but this was

qualified by a significant interaction, $F(1, 105) = 5.45, p = .021, \eta_p^2 = .05$. Again, participants in the longhand-study condition (z-score $M = 0.68$) outperformed the other participants (z-score $M_s = -0.09, -0.28, -0.34$), $t(105) = 4.50, p < .001, d = 0.88$. Among students who had the opportunity to study, longhand note takers did significantly better than laptop note takers, $t(53) = 2.77, p = .008$.

Collapsing performance on conceptual, inferential, and application questions into a general “conceptual” measure revealed no significant main effects, but again there was a significant interaction between note-taking medium and studying. There was a significant interaction for conceptual questions, $F(1, 105) = 4.35, p = .04, \eta_p^2 = .04$. Among students who had the opportunity to study, longhand note takers did significantly better than laptop note takers, $t(53) = 2.32, p = .024, d = 0.64$.

Table 2. Raw Means for Overall, Factual, and Conceptual Performance in the Four Conditions of Study 3

Question type	Longhand-study	Longhand-no study	Laptop-study	Laptop-no study
Factual only	7.1 (4.0)	3.8 (2.8)	4.5 (3.2)	3.7 (3.1)
Conceptual only	18.5 (7.8)	15.6 (7.8)	13.8 (6.3)	16.9 (8.1)
Overall	25.6 (10.8)	19.4 (9.9)	18.3 (9.0)	20.6 (10.7)

Note: Standard deviations are given in parentheses. Although raw scores are given here, z scores of the index scores were used in the analysis.

Content analysis of notes. Again, longhand note takers wrote significantly fewer words ($M = 390.65$, $SD = 143.89$) than those who typed ($M = 548.73$, $SD = 252.68$), $t(107) = 4.00$, $p < .001$, $d = 0.77$ (see Fig. 2). As in the previous studies, there was a significant difference in verbatim overlap, with a mean of 11.6% overlap ($SD = 5.7\%$) for laptop note taking and only 4.2% ($SD = 2.5\%$) for longhand, $t(107) = 8.80$, $p < .001$, $d = 1.68$ (see Fig. 3). There were no significant differences in word count or verbatim overlap between the study and no-study conditions.

The amount of notes taken positively predicted performance for all participants, $\beta = 0.35$, $p < .001$, $R^2 = .12$. The extent of verbatim overlap did not significantly predict performance for participants who did not study their notes, $\beta = 0.13$. However, for participants who studied their notes (and thus those who were most likely to be affected by the contents), verbatim overlap negatively predicted overall performance, $\beta = -0.27$, $p = .046$, $R^2 = .07$. When looking at overall test performance, longhand note taking negatively predicted word count, which positively predicted performance, indirect effect = -0.15 , 95% CI = $[-0.24, -0.08]$. Longhand note taking also negatively predicted verbatim overlap, which negatively predicted performance, indirect effect = 0.096 , 95% CI = $[0.004, 0.23]$.

However, a more nuanced story can be told; the indirect effects differ for conceptual and factual questions. For conceptual questions, there were significant indirect effects on performance via both word count (-0.17 , 95% CI = $[-0.29, -0.08]$) and verbatim overlap (0.13 , 95% CI = $[0.02, 0.15]$). The indirect effect of word count for factual questions was similar (-0.11 , 95% CI = $[-0.21, -0.06]$), but there was no significant indirect effect of verbatim overlap (0.04 , 95% CI = $[-0.07, 0.16]$). Indeed, for factual questions, there was no significant direct effect of overlap on performance ($p = .52$). As in Studies 1 and 2, the detriments caused by verbatim overlap occurred primarily for conceptual rather than for factual information, which aligns with previous literature showing that verbatim note taking is more problematic for conceptual items (e.g., Bretzing & Kulhavy, 1979).

When participants were unable to study, we did not see a difference between laptop and longhand note taking. We believe this is due to the difficulty of test items after a week's delay and a subsequent floor effect; average scores were about one-third of the total points available. However, when participants had an opportunity to study, longhand notes again led to superior performance. This is suggestive evidence that longhand notes may have superior external-storage as well as superior encoding functions, despite the fact that the quantity of notes was a strong positive predictor of performance. However, it is also possible that, because of enhanced encoding, reviewing longhand notes simply reminded participants of lecture information more effectively than reviewing laptop notes did.¹⁰

General Discussion

Laptop note taking has been rapidly increasing in prevalence across college campuses (e.g., Fried, 2008). Whereas previous studies have shown that laptops (especially with access to the Internet) can distract students, the present studies are the first to show detriments due to differences in note-taking behavior. On multiple college campuses, using both immediate and delayed testing across several content areas, we found that participants using laptops were more inclined to take verbatim notes than participants who wrote longhand, thus hurting learning. Moreover, we found that this pattern of results was resistant to a simple verbal intervention: Telling students not to take notes verbatim did not prevent this deleterious behavior.

One might think that the detriments to encoding would be partially offset by the fact that verbatim transcription would leave a more complete record for external storage, which would allow for better studying from those notes. However, we found the opposite—even when allowed to review notes after a week's delay, participants who had taken notes with laptops performed worse on tests of both factual content and conceptual understanding, relative to participants who had taken notes longhand.

We found no difference in performance on factual questions in the first two studies, though we do not discount the possibility that with greater power, differences might be seen. In Study 3, it is unclear why longhand note takers outperformed laptop note takers on factual questions, as this difference was not related to the relative lack of verbatim overlap in longhand notes. It may be that longhand note takers engage in more processing than laptop note takers, thus selecting more important information to include in their notes, which enables them to study this content more efficiently. It is worth noting that longhand note takers' advantage on retention of factual content is limited to conditions in which there was a delay between presentation and test, which may explain the discrepancy between our studies and previous research (Bui et al., 2013). The tasks they describe would also fall under our factual-question category, and we found no difference in performance on factual questions in immediate testing. For conceptual items, however, our findings strongly suggest the opposite conclusion. Additionally, whereas Bui et al. (2013) argue that verbatim notes are superior, they did not report the extent of verbatim overlap, merely the number of "idea units." Our findings concur with theirs in that more notes (and therefore more ideas) led to better performance.

The studies we report here show that laptop use can negatively affect performance on educational assessments, even—or perhaps especially—when the computer is used for its intended function of easier note taking.

Although more notes are beneficial, at least to a point, if the notes are taken indiscriminately or by mindlessly transcribing content, as is more likely the case on a laptop than when notes are taken longhand, the benefit disappears. Indeed, synthesizing and summarizing content rather than verbatim transcription can serve as a desirable difficulty toward improved educational outcomes (e.g., Diemand-Yauman, Oppenheimer, & Vaughan, 2011; Richland, Bjork, Finley, & Linn, 2005). For that reason, laptop use in classrooms should be viewed with a healthy dose of caution; despite their growing popularity, laptops may be doing more harm in classrooms than good.

Author Contributions

Both authors developed the study concept and design. Data collection was supervised by both authors. P. A. Mueller analyzed the data under the supervision of D. M. Oppenheimer. P. A. Mueller drafted the manuscript, and D. M. Oppenheimer revised the manuscript. Both authors approved the final version for submission.

Acknowledgments

Thanks to Jesse Chandler, David Mackenzie, Peter Mende-Siedlecki, Daniel Ames, Izzy Gainsburg, Jill Hackett, Mariam Hambarchyan, and Katelyn Wirtz for their assistance.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Supplemental Material

Additional supporting information may be found at <http://pss.sagepub.com/content/by/supplemental-data>

Open Practices



All data and materials have been made publicly available via Open Science Framework and can be accessed at <http://osf.io/crsiz>. The complete Open Practices Disclosure for this article can be found at <http://pss.sagepub.com/content/by/supplemental-data>. This article has received badges for Open Data and Open Materials. More information about the Open Practices badges can be found at <https://osf.io/tvyxz/wiki/view/> and <http://pss.sagepub.com/content/25/1/3.full>.

Notes

1. See Additional Analyses in the Supplemental Material available online for some findings regarding real-world data.
2. See Lecture Information in the Supplemental Material for links to all five TED Talks used in Study 1 and the four prose passages used in Study 2.
3. See Raw Means and Questions in the Supplemental Material for full question lists from all three studies.
4. For factual questions, laptop participants' raw mean score was 5.58 ($SD = 2.23$), and longhand participants' raw mean

score was 6.41 ($SD = 2.84$). For conceptual questions, the raw mean scores for laptop and longhand participants were 3.77 ($SD = 1.23$) and 4.29 ($SD = 1.49$), respectively. See Raw Means and Questions in the Supplemental Material for raw means from Studies 1 and 2.

5. In all three studies, the results remained significant when we controlled for measures of academic ability, such as self-ratings of prior knowledge and scores on the SAT and reading span task.
6. Linguistic Inquiry and Word Count (LIWC) software was also used to analyze the notes on categories identified by Pennebaker (2011) as correlating with improved college grades. Although LIWC analysis indicated significant differences in the predicted direction between laptop and longhand notes, none of the differences predicted performance, so they will not be discussed here.
7. For all three studies, we also analyzed the relation between verbatim overlap and students' preferences for longhand or laptop note taking. Results of these analyses can be found in Additional Analyses in the Supplemental Material.
8. For conceptual questions, laptop-nonintervention participants had lower raw scores ($M = 2.30$, $SD = 1.40$) than did longhand note takers ($M = 2.94$, $SD = 1.73$) and laptop-intervention participants ($M = 2.43$, $SD = 1.59$). For factual questions, laptop-nonintervention participants' raw scores ($M = 4.92$, $SD = 2.62$) were also lower than those of longhand note takers ($M = 5.11$, $SD = 3.05$) or laptop-intervention participants ($M = 5.25$, $SD = 2.89$).
9. Estimating effect sizes in mixed models is problematic (for more details, see, e.g., Judd, Westfall, & Kenny, 2017). Thus, while the η_p^2 s reported above for Studies 1 and 2 (ranging from .66 to .95) are what SPSS reports, more informative estimates of effect sizes would be much smaller, ranging from .03 to .11.
10. Participants who took laptop or longhand notes but who did not have the opportunity to study did not score significantly differently on either factual, $t(52) = 0.26$, $p = .795$, $d = 0.07$, or conceptual, $t(52) = 0.77$, $p = .442$, $d = 0.21$, questions.

References

- Aiken, E. G., Thomas, G. S., & Shennum, W. A. (1975). Memory for a lecture: Effects of notes, lecture rate, and informational density. *Journal of Educational Psychology*, 67, 439–444.
- Barak, M., Lipson, A., & Lerman, S. (2006). Wireless laptops as means for promoting active learning in large lecture halls. *Journal of Research on Technology in Education*, 38, 245–263.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128, 612–637.
- Bohay, M., Blakely, D. P., Tamplin, A. K., & Radvansky, G. A. (2011). Note taking, review, memory, and comprehension. *The American Journal of Psychology*, 124, 63–73.
- Bretzing, B. H., & Kulhavy, R. W. (1979). Notetaking and depth of processing. *Contemporary Educational Psychology*, 4, 145–153.
- Bretzing, B. H., & Kulhavy, R. W. (1981). Note-taking and passage style. *Journal of Educational Psychology*, 73, 242–250.
- Brown, C. M. (1988). Comparison of typing and handwriting in "two-finger typists." *Proceedings of the Human Factors Society*, 32, 381–385.
- Bui, D. C., Myerson, J., & Hale, S. (2013). Note-taking with computers: Exploring alternative strategies for improved recall. *Journal of Educational Psychology*, 105, 299–309.

- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1118–1133.
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*, 116–131.
- Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.
- Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. *Cognition*, *118*, 111–115.
- DiVesta, F. J., & Gray, G. S. (1972). Listening and note taking. *Journal of Educational Psychology*, *63*, 8–14.
- DiVesta, F. J., & Gray, S. G. (1973). Listening and note taking: II. Immediate and delayed recall as functions of variations in thematic continuity, note taking, and length of listening-review intervals. *Journal of Educational Psychology*, *64*, 278–287.
- Einstein, G. O., Morris, J., & Smith, S. (1985). Note-taking, individual differences, and memory for lecture information. *Journal of Educational Psychology*, *77*, 522–532.
- Fried, C. B. (2008). In-class laptop use and its effects on student learning. *Computers & Education*, *50*, 906–914.
- Garner, R., Gillingham, M. G., & White, C. S. (1989). Effects of “seductive details” on macroprocessing and microprocessing in adults and children. *Cognition and Instruction*, *6*, 41–57.
- Grace-Martin, M., & Gay, G. (2001). Web browsing, mobile computing and academic performance. *Educational Technology & Society*, *4*, 95–107.
- Hembroke, H., & Gay, G. (2003). The laptop and the lecture: The effects of multitasking in learning environments. *Journal of Computing in Higher Education*, *15*, 46–64.
- Igo, L. B., Bruning, R., & McCrudden, M. T. (2005). Exploring differences in students’ copy-and-paste decision making and processing: A mixed-methods study. *Journal of Educational Psychology*, *97*, 103–116.
- Iqbal, S. T., & Horvitz, E. (2007). Disruption and recovery of computing tasks: Field study, analysis, and directions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 677–686). New York, NY: Association for Computing Machinery. doi:10.1145/1240624.1240730
- Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, *68*, 601–625.
- Kay, R., & Lauricella, S. (2011). Exploring the benefits and challenges of using laptop computers in higher education classrooms: A formative analysis. *Canadian Journal of Learning and Technology*, *37*(1). Retrieved from <http://www.cjlt.ca/index.php/cjlt/article/view/565/299>
- Kiewra, K. A. (1985). Investigating notetaking and review: A depth of processing alternative. *Educational Psychologist*, *20*, 23–32.
- Kiewra, K. A. (1989). A review of note-taking: The encoding-storage paradigm and beyond. *Educational Psychology Review*, *1*, 147–172.
- Kobayashi, K. (2005). What limits the encoding effect of note-taking? A meta-analytic examination. *Contemporary Educational Psychology*, *30*, 242–262.
- Kraushaar, J. M., & Novak, D. C. (2010). Examining the affects of student multitasking with laptops during the lecture. *Journal of Information Systems Education*, *21*, 241–251.
- Lin, L., & Bigenho, C. (2011). Note-taking and memory in different media environments. *Computers in the Schools*, *28*, 200–216.
- Mitra, A., & Steffensmeier, T. (2000). Changes in student attitudes and student computer use in a computer-enriched environment. *Journal of Research on Computing in Education*, *32*, 417–432.
- Pennebaker, J. W. (2011). *The secret life of pronouns: What our words say about us*. New York, NY: Bloomsbury Press.
- Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, and Computers*, *36*, 717–731.
- Quade, A. Q. (1996). An assessment of retention and depth of processing associated with notetaking using traditional pencil and paper and an on-line notepad during computer-delivered instruction. In M. R. Simonson, M. Hays, & S. Hall (Eds.), *Proceedings of Selected Research and Development Presentations at the 1996 Convention of the Association for Educational Communications and Technology* (pp. 559–570). Retrieved from <http://files.eric.ed.gov/fulltext/ED397772.pdf>
- Richland, L. E., Bjork, R. A., Finley, J. R., & Linn, M. C. (2005). Linking cognitive science to education: Generation and interleaving effects. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society* (pp. 1850–1855). Mahwah, NJ: Erlbaum.
- Rubinstein, J. S., Meyer, D. E., & Evans, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 763–797.
- Skolnick, R., & Puzo, M. (2008). Utilization of laptop computers in the school of business classroom. *Academy of Educational Leadership Journal*, *12*, 1–10.
- Slotte, V., & Lonka, K. (1999). Review and process effects of spontaneous note-taking on text comprehensions. *Contemporary Educational Psychology*, *24*, 1–20.
- Sovern, J. (2013). Law student laptop use during class for non-class purposes: Temptation v. incentives. *University of Louisville Law Review*, *51*, 483–517.
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505.
- Van Meter, P., Yokoi, L., & Pressley, M. (1994). College students’ theory of note-taking derived from their perceptions of note-taking. *Journal of Educational Psychology*, *86*, 325–338.
- Wurst, C., Smarkola, C., & Gaffney, M. A. (2008). Ubiquitous laptop usage in higher education: Effects on student achievement, student satisfaction, and constructivist measures in honors and traditional classrooms. *Computers & Education*, *51*, 1766–1783.
- Yamamoto, K. (2007). Banning laptops in the classroom: Is it worth the hassle? *Journal of Legal Education*, *57*, 477–520.