

Selection-Based Note-Taking Applications

Aaron Bauer, Kenneth R. Koedinger
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA 15213
{abauer, koedinger}@cmu.edu

ABSTRACT

The increasing integration of education and technology has led to the development of a range of note-taking applications. Our project's goal is to provide empirical data to guide the design of such note-taking applications by evaluating the behavioral and learning outcomes of different note-taking functionality. The study reported here compares note-taking using a text editor and four interaction techniques. The two standard techniques are typing and copy-paste. The two novel techniques are restricted copy-paste and menu-selection, intended to increase attention and processing respectively. Hypothesized learning gains from the novel techniques were not observed. As implemented these techniques were less efficient and appeared to be more frustrating to use. However, data regarding differences in both note-taking efficiency and learning suggest several important implications for selection-based note-taking applications, such as pasting and highlighting. Our results also indicate that students have strong opinions regarding their note-taking practices, which may complicate potentially beneficial interventions.

Author Keywords

Note-taking, education, annotation.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI):
Miscellaneous.K.3 Computers and Education

INTRODUCTION

Note-taking is an important student behavior. Studies have shown that as many as ninety-nine percent of college students take notes in lecture [14], and ninety-four percent of college students believe note-taking is an important part of the educational experience [17]. As more of our educational material moves to the computer, supporting note-taking digitally becomes an important task. At the same time, technology gives us unprecedented control over

the note-taking process. Our research is aimed at using this control to study note-taking processes in greater detail. In doing so, we hope to provide developers with guidelines for creating note-taking tools that encourage learning. As of yet, this is an unrealized opportunity. Though a few applications have shown changes in note-taking behavior, none have systematically investigated the effect of these changes on learning.

In this paper, we report the results of an experiment evaluating the impact of several selection-based note-taking tools on both student behavior and learning. We also describe potential implications of our results for selection-based note-taking applications such as digital highlighting.

PREVIOUS TECHNOLOGY

Designers of note-taking applications must make decisions regarding which features to include. Even a simple tool such as a text-editor embedded in online course content requires that decisions be made. Should copy-paste functionality be included? There are arguments in both directions. Such a feature would likely make note-taking faster and more efficient. However, it may encourage behaviors that reduce learning by decreasing attention and increasing verbatim note-taking. Unfortunately, empirical data does not exist to help designers make even this decision. Providing this type of data to designers is the goal of our research.

Many projects involving in-class and online educational technology include note-taking applications. Some of these applications have even been evaluated with regards to note-taking behavior. These evaluations indicate that technology can have a strong impact on note-taking. However, few have studied the impact these changes have on learning, and none in a systematic fashion.

In the absence of empirical data to drive design, note-taking applications generally either mimic traditional note-taking practices or support new methods of note-taking. Online note-taking systems often mimic real world annotation of paper documents (as described in [11]). An example of a full-featured note-taking device of this type is Sidenote, a commercial application that supports most styles of annotation, including drawing, highlighting, and text-entry [10].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2007, April 28–May 3, 2007, San Jose, California, USA.

Copyright 2007 ACM 978-1-59593-593-9/07/0004...\$5.00.

Few of these devices have been evaluated. However, restrictions placed on annotation have been shown to change the ways students take notes. For example, when students can only attach comments to highlighted areas, they may find it difficult to make general summary comments that would otherwise have been written as marginal notes [12].

Stupad provides an interesting example of note-taking interventions that change note-taking behavior in predictable ways. Stupad was initially designed to allow students to take freeform notes during lecture [15], mimicking pencil and paper. After class, notes were attached to audio and content streams from the lecture. The second iteration of Stupad integrated live captures of content streams in order to overcome the perceived problem of students recording redundant notes (notes already being captured in the content stream) verbatim. The new version functioned as intended, reducing verbatim note-taking, increasing summary notes, and reducing overall note-taking.

Many novel note-taking applications are developed to support note-sharing. For example, anchored discussion applications allow students to share and discuss their annotations of learning material [e.g. 12]. Note-sharing functionality can also have an impact on how students take notes. One study found that a note sharing task changed both online and offline note-taking. Students reported foregoing their personal wording in order to increase readability. To facilitate the task of sharing digital annotations, some even changed their note-taking style when annotating paper printouts [12].

The Livenotes project [8] is an example of another application supporting novel note-taking procedures. Small groups of students use the Livenotes system to jointly take notes on top of slides in a virtual whiteboard. This new style of note-taking, based on theories of collaborative learning which suggest that interaction promotes learning, eliminates personal notes. Once again students report that they change their note-taking style to make their notes more comprehensible to others. In addition, when compared to students using the Livenotes system individually, group notes are shown to have far more comments and question and answer threads.

These devices demonstrate the effect of functionality on note-taking, but give no indication of resulting learning differences. Therefore, they give few hints as to which capabilities should be included in a note-taking application intended to promote learning. There are many features a designer could support, from highlighting to text-entry to freeform annotation, but little empirical direction. Though it is clear we can manipulate note-taking behavior, the educational value of the changes in behavior is unclear. Investigating effect of digitally supported note-taking on

learning could help drive the design of tools that support learning.

NOTE-TAKING RESEARCH

Note-taking has been shown to promote learning. Researchers generally propose two potential benefits of note-taking: encoding and artifact. The encoding hypothesis is the idea that the act of note-taking facilitates learning. Note-taking requires students to identify important material, which they then transcribe into their own notes. This increases processing of the material, augmenting learning. The external artifact hypothesis states that having notes for review facilitates learning: notes aid students most when they can be reviewed.

Note-taking studies have demonstrated positive encoding and external artifact effects, though it appears that the external artifact effect may be more reliable. In a review by Kiewra, slightly more than half of encoding studies found when note-taking was compared with not taking notes. Three-fourths of external artifact studies found positive effects [9].

Our early studies [1] indicated that encoding effects were most likely to be effected by our interventions. Though as mentioned above, encoding is an established benefit of note-taking, its foundations are unclear. Two of the most common hypotheses regard the efficiency and wording with which notes are recorded. Previous research suggests that both can be influenced, and thus studied, using note-taking technology.

Efficiency

Efficient notes can be described with regards to the number of ideas recorded, or the wordiness with which each idea is recorded. Tests of recall (free-response or essay) have found that ideas present in notes are more likely to be recalled [4, 6] than ideas absent from notes. However, efficiently recording ideas may involve focusing on the key ideas, and ignoring extra or irrelevant details. In this view, the ratio of key ideas to total ideas recorded is the important metric of efficiency. Research has not been done on this topic.

The desirability of wordy notes has been studied, but the results are equivocal. Some studies indicate that more “efficient notes”, operationalized as lower ratios of words to ideas, are correlated with successful learning outcomes [6]. Others failed to confirm this result [5]. Still others have found that an increased number of words in notes correlates positively with learning outcomes [13].

Our earlier studies have shown that both of these measures of efficiency are affected by the features included in note-taking technology. In particular, students given copy-paste functionality appear to increase the total number of ideas recorded as well as the wordiness with which each idea is expressed [1]. Understanding the impact (if any) of these

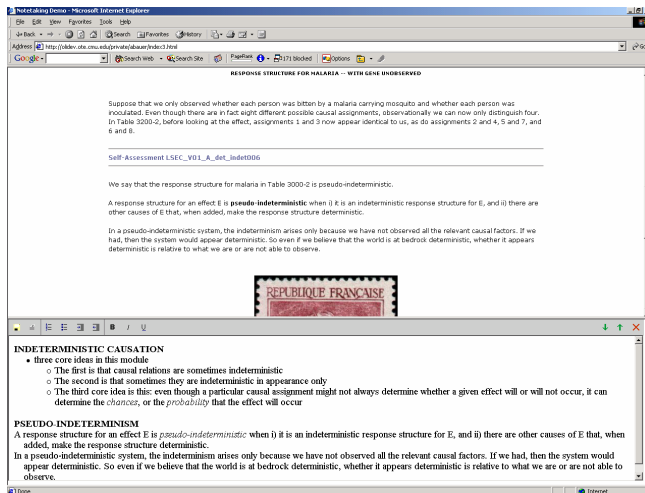


Figure 1: Note-Taking Interface

changes on learning is important in determining whether such a feature should be included in a note-taking device.

Wording

Many researchers believe it is important that students use their own words to record notes. However, verbatim notes are common. One study found that students preparing for a presentation took notes that were 60% verbatim [3]. The Stupad project showed we can manipulate the amount of verbatim notes recorded using a note-taking device [15]. Some researchers believe that verbatim notes, as opposed to notes recorded in students' own words, are not conducive to learning, as they allow students to record material without processing it very deeply. Pasting functionality could further reduce processing of verbatim notes, as it requires even less attention and time on task than handwriting or typing.

There are studies that examine the impact of verbatim notes on learning. In one, students who were only allowed to take verbatim notes while reading showed poorer outcomes on both immediate and delayed tests when compared with students who took paraphrased or summarized notes [2]. However, students report strategies for taking verbatim notes, finding them most useful when recording facts (as opposed to ideas), definitions, and when they believe verbatim transcriptions will help recognize material on tests [16]. Being forced to take verbatim notes may have reduced the effectiveness of such strategies.

OUR RESEARCH

Our project attempts to provide developers of note-taking applications with empirical data that can be used to make decisions regarding functionality. We aim to do so by investigating how note-taking interaction techniques influences learning. Our applications are being developed for an interdisciplinary program that is creating online courses in areas such as Logic, Physics and Chemistry. For

several reasons, we have started with a basic embedded JavaScript text-editor (see Figure 1), rather than one of the more advanced applications described above.

First, most prior literature on note-taking involves notes students write on paper. As we would like to expand upon and derive lessons from these results, we wanted to start with a tool that allowed behaviors most similar to this traditional style. Text-editors, in which students may create notes detached from learning materials, are a better match than devices that annotate the material itself.

We also believe that results from studies of text-editing can transfer to annotation applications much more easily than can occur in the opposite direction. If we offered a multi-featured annotation application, it would be difficult to ascribe any differential learning outcomes to specific features, reducing generalizability. A more limited annotation application, perhaps only allowing highlighting, would limit the types of notes we would see from students. A text-editor can be used to evaluate selection techniques, such as copy-paste, while still allowing us to study content students create themselves by typing.

In a previous study [1], we evaluated the impact of copy-paste functionality on note-taking. Students took notes using either pencil-and-paper, a text editor that allowed copy-paste, or a text-editor that prohibited copy-paste. Learning was evaluated at three intervals: immediately following study, at a week's delay, and following a short review period after the delayed test. We found that students in the Paste treatment took advantage of the paste facility to record far more ideas, of a far more wordy and verbatim nature. Though there was no difference on individual tests or in the amount of time spent on learning material, students in the Paste treatment appeared to forget significantly more over time than did students in the other treatments. There was also an interesting contrast of the wordiness of ideas recorded; whereas for the Paste treatment wordy notes resulted in more forgetting, the opposite was seen for the typing and Paper treatments. We did not find any effects for how the notes were recorded, as notes recorded verbatim appeared to perform equivalently to notes recorded in students own words. It is important to note that while students in the Paste treatment were allowed to type, many did not.

Hypotheses

We developed two hypotheses regarding the root cause of the copy-paste effect, and designed interventions to test these hypotheses.

Typing and handwriting take more time and effort than pasting. In order to reduce the amount of material to transcribe, students must make decisions regarding what is most important to transcribe. It is possible that this leads them to focus in on critical details, as indicated by the reduced wordiness in these two treatments. This focus on

critical information may be responsible for increased encoding.

To test this, we designed a note-taking application that allowed students to paste, but restricted the amount of text they could paste in any single action in order to focus their attention on the critical information. We explored two alternative designs. The first allowed students to select whatever they wanted. When the selection was too long it was automatically deselected. The second design created hard stops at selection boundaries, so that students could not make lengthy selections. We decided on the former, as we were concerned the latter would allow the student to select without paying attention, or that students would not notice that the final selection was not what they had intended. User of this interface tests produced the intended behavioral results. Further details of the boundary definitions are provided below.

One previous study of copy-paste restriction showed learning gains realized by restricted note-taking [7]. However, in this note-taking treatment students were given tables to fill out using either restricted or unrestricted copy-paste. Therefore the results may not generalize for several reasons. First, it requires highly organized learning material, which can be categorized hierarchically. In addition to requiring extra work from designers to create the table, the table headings themselves provide extra information to students, giving effective note-taking hints to students. Here we test selection techniques that are more student-motivated and less organized.

Typing and handwriting allow students to do two things copy-paste do not. First, they create their own notes. Secondly, they have the ability to reword notes. It may be that the generation of notes increases encoding, or that viewing an additional representation (i.e. rewording) of a concept increases encoding. To test this question, we designed an application that, when a selection was made in text, gives the student three options to place in their notes, one of which was a reworded version of the concept (see Figure 2). If creation was important, this would show superior learning to the paste tool. If the rewording was important, this would show similar performance to the Paste tool.

For experimental purposes, the Select tool was not designed to give feedback when the user made an incorrect selection, as the users of other tools do not receive feedback. However, as a practical device, the selection tool may be attractive in part for its ability to give feedback. Instructors could target misconceptions and give self-assessment questions while at the same time allowing students to complete note-taking tasks.

STUDY

Design and Subjects

The study presented here compares students taking notes in an online classroom using four treatments: “Paste”, “Restricted Paste”, “Typing”, and “Selection”. It followed a between-subjects design, so participants were randomly assigned to one note-taking method. Three learning measures are obtained: immediate, delayed, and delayed with review.

The “Typing” treatment remained the same as in the previous study. Students could only type information into the notepad. The tool did not give alerts if students attempted to paste or drag contents into their notes; it simply did not let them.

In order to avoid a potential self-selection bias present in the last study, where students in the Paste treatment could type when pasting would be most detrimental to learning, we eliminated the ability of the Paste treatment to type. We expected this to reduce the effectiveness of the Paste treatment with regards to learning.

The Restricted treatment was identical to the Paste treatment, except students could only select a limited amount of text in any one action. This restriction was identified as only 90% of the words of any given sentence. When a student made an illegal selection, it was automatically deselected by the application. Selections could not cross sentence-boundaries. No feedback was given, though these behaviors were explained to students beforehand.

In the Select treatment, whenever students made a selection in text, a box with three options became available next to the cursor. Two of the options were distractors, while the third was a reworded representation of the selected concept (see Figure 2).

A total of 76 subjects from several local universities were recruited by means of a posting to a subject-recruitment website. Two students did not show up for the second day, and 3 who were given incorrect quiz materials. Their data was not included in the analyses described here. No students reported being familiar with the course materials, and none employed a “hunt-and-peck” typing strategy. Participants were paid per hour participated.

Materials

Participants completed one module in an online course in Causal and Statistical Reasoning. The module is 13 pages long and consists of approximately 9000 words. In our previous experiment, interactive simulations and comprehension questions were available. In this experiment, these were not included in order to increase the focus on note-taking.

be understood as relative to a particular set of background conditions. For example, consider the claim: "In America, the more calories you eat, the more likely you are to be depressed." Causal generalizations are relative to the Response Structure. At certain points, causal generalizations are relative to the background conditions. At other points, causal generalizations are relative to the interactions among variables. The Homenet Study

es on the effect of spending time on the internet. In 1995, the Homenet Project at Carnegie Mellon found that th ore likely they were to be depressed. A repeat of the study in 2000 found just the opposite. What was the expla from 1995 to 2000. In 1995, approximately 20% of the population was internet enabled, in 2000 the number was ing online to strangers leads to depression, but talking online to family and friends reduces it. The more time so ng to strangers. In 2000, the more they were talking to family and friends.

Figure 2: Selection Tool. Selecting entry places it in text editor below the content.

Participants were seated at a desk in front of a 17-inch monitor whose resolution was set at 1024 by 768 pixels, a keyboard, and a mouse. The course content took the top third of the screen (see Figure 1), with the text-editor in the bottom third. Students could modify this ratio, and our survey did indicate a level of dissatisfaction with the screen size. The JavaScript-based text-editor supported basic markup: bold, italic, and underline. It also supported basic outlining: indenting, bulleted lists, and ordered lists. Keyboard shortcuts were available for markup, as well as pasting for participants in the Paste treatment. Students in all treatments could copy-paste within the notepad, and type dashes, parentheses and similar outlining characters.

Each quiz contained 25 items, which targeted the 10 ideas around which the instructor based the module and the three 12 item multiple choice test which were the basis of our quizzes. Though questions only differed with regards to context, not format, we did not have data to match them statistically. Therefore we completely counterbalanced the presentation of the tests, so that in each treatment some would start with test A, others would start with test B, and the rest would start with test C. The tests had 11 multiple-choice items and 14 free response items.

Procedure

After informed consent was obtained, participants were told that they would be studying the second module in an online course in Causal and Statistical Reasoning, and were given a one minute typing speed test. Upon its completion they were asked to read a one-page summary of the introductory module. This was followed by a pretest, identical in form to the learning tests. They were then introduced to the task. Participants were asked to take notes while studying, and told that their notes would be available for review during the second session. Students were told to use their assigned tool to record any notes they would want to review the following week. Participants were allowed as much time as they required to complete the module. Immediately after completing the module, participants were given the first test to complete. This was the final activity on the first day.

The second session was conducted seven days later. Participants were given the delayed test when they arrived. Upon its completion, participants were provided with their notes, and told they would have five minutes for review.

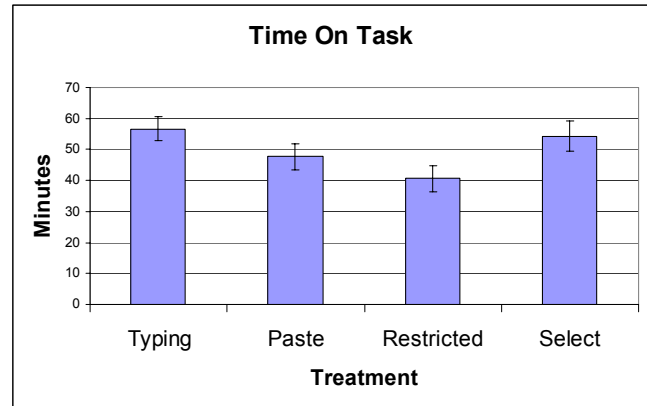


Figure 3: Time On Task

They were required to use the full five minutes, and instructed to review mentally if they finished reviewing their notes before the five minutes were completed. They were then given the final test. After the test they were given a short survey regarding their experience.

Dependent Measures

We collected several dependent measures, including completion time, measures of notes, test scores. Notes were coded with regards to total words, as well as individual idea units. Notes were first split by HTML paragraphs, which were then further segmented into individual sentences. These were then matched with ideas present in the notes, which required that some sentences be further divided. The majority of idea units were identified by matching phrasing within the note with verbatim phrasing within the module. The few that remained were identified by the page on which they were recorded and neighboring notes.¹

In addition to the distinction between Multiple Choice and Free Response items, there is a distinction between test items that require students to recite information and test items that require students to use information to answer questions. Recite items are all in Free Response form.

RESULTS

We performed an ANOVA on time on task. SAT was included as a covariate in the model (pretest was not significant), in a full factorial with treatment. There was a significant treatment effect of time taken to study ($F(3,58)=3.4, p=.02$) (see Figure 3). Contrasts showed that Typing took significantly longer than pasting ($p=.04$) and

¹As only the typing tool allowed non-verbatim ideas, most ideas could be identified by simple searches. Most Typed ideas could be identified in a similar fashion. Even so, although coding was blind to treatment, a weakness of this study is that all ideas were identified by one researcher.

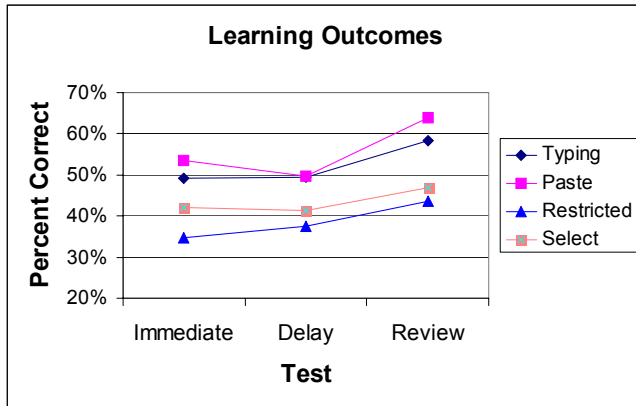


Figure 4: Overall Learning Results

Restricted ($p=.003$) The contrast between Select and Restricted was marginal ($p<.1$).

Learning

In our learning analyses, ANCOVA was performed with item correctness as the dependent measure, pre-test as a covariate, treatment as a between-subjects variable (Paste-only vs. Typing vs. Select vs. Restricted) and both test-time (immediate vs. delay vs. review) and item-type (Multiple-Choice vs. Free-Response) as within-subjects variables in a full factorial. Subject was included as a random effect, as each subject answered many questions per test. There were main effects of treatment [$F(3,66)=2.9, p <.05$], pre-test [$F(1,5369)=40, p<.0001$], test-time [$F(2,66)=17.7, p<.0001$], and item-type [$F(1,5369)=6.86, p<.01$]. The only marginally significant interaction was test-time by item-type [$F(6,5369)=2.58, p=.07$].

As seen in Figure 4, students who used the novel tools (Select and Restricted) learned less than those using the less novel tools (Typing and Paste). Contrasts between treatments showed several marginal and significant results. Paste was marginally superior to Select ($p=.1$) and

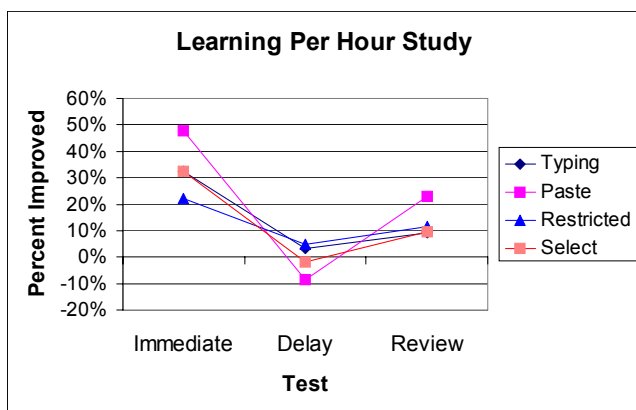


Figure 5: Learning Efficiency

significantly better than Restricted ($p=.007$). Typing was also significantly better than Restricted ($p=.03$). Redoing the above ANCOVA with the treatments grouped as more-novel vs. less-novel yields a significant novelty effect [$F(1,66)=7.1, p<.01$].

Investigating the test-time effect shows no Forgetting effect between the immediate and delayed test [$F(1,5369)=.03, p=.84$], but does show a significant Review effect [$F(1,5369)=27.5, p<.0001$]. However, as mentioned above there was not a significant treatment by test-time interaction, indicating that there were no treatment differences in how much was forgotten between the immediate and delayed test or relearned between the delayed test and the review test. The item-type effect is indicative of superior performance on multiple-choice items.

As treatment appears to affect time on task, we conducted an evaluation of learning efficiency, which indicated efficiency effects in favor of the Paste treatment (see Figure 5). This measure was operationalized as the difference between two in-sequence tests (i.e., pretest and immediate or delay and review) divided by study time (i.e. “time on task”). Our analyses indicate Paste was the most efficient tool. As was the case with time on task, SAT Math was a significant covariate, whereas pretest was not. ANOVAS were conducted with Treatment and SAT-Math in the model. The overall efficiency effect was marginal for the immediate test [$F(3,58)=2.34, p=.08$] and the review test [$F(3,58)=1.9, .12$], and not different for the delayed test [$F(3,58)=1.3, p=.27$]. On the immediate test, Paste was significantly better than Restricted ($p=.01$), and marginally better than Typing ($p=.11$) and Selection (.15). Paste was significantly better than Typing on Review ($p=.03$), and marginally better than both Selection (.07) and Restricted (.09). Contrasts between Typing, Restricted, and Selection were never significant.

Note-Taking

ANOVAs were conducted on each note-taking measure, with treatment (Paste-only vs. Typing vs. Select vs. Restricted), and SAT-Math included in the model in a full factorial. Neither Typing Speed nor Pretest were included in the model, as they were not found to be significant covariates. SAT-Math was included as it was significant for all measures described below but Wordiness, where it is marginal for Ideas ($p=.09$) and not significant for key ideas.

There is a significant effect found for treatment with regards to both total words [$F(3,58)=8.9, p<.0001$] and total ideas recorded [$F(3,58)=8.0, p=.0001$]. Word treatment contrasts show that paste is significantly different from all other treatments (all $p<.001$), none of which are significantly different from each other. The same is true for ideas (see Figure 6), though now the Typing-Restricted contrast is marginal ($p=.06$). There is an overall wordiness

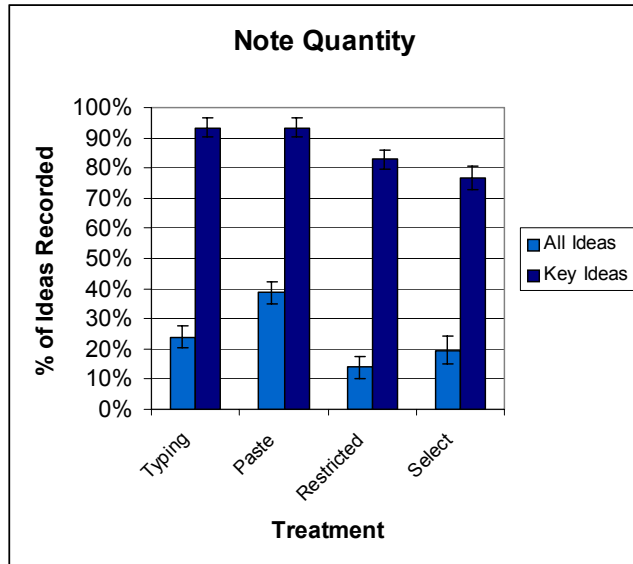


Figure 6: Note Quantity: For display purposes, “All Ideas” are divided by the greatest number of ideas recorded, and “Key Ideas” are divided by the total number of key ideas in the module.

effect as well [$F(3,58)=8.0, p=.0001$], where Paste is more wordy than all other tools. Though there is a significant Restricted-Select contrast with regards to wordiness ($p=.03$), it should be noted that the Select treatment did not have control over wording. There is no significant difference between the Restricted and Typing treatments.

Key ideas represent the 10 ideas around which the learning contents and test-items are designed. There is a significant effect for key ideas recorded [$F(3,58)=5.9, p=.001$]. In this case, the “novel” tools (Restricted and Select) record

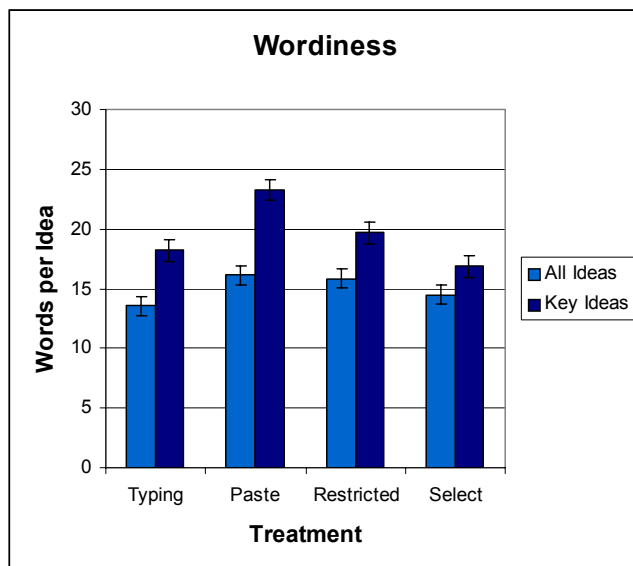


Figure 7: Wordiness

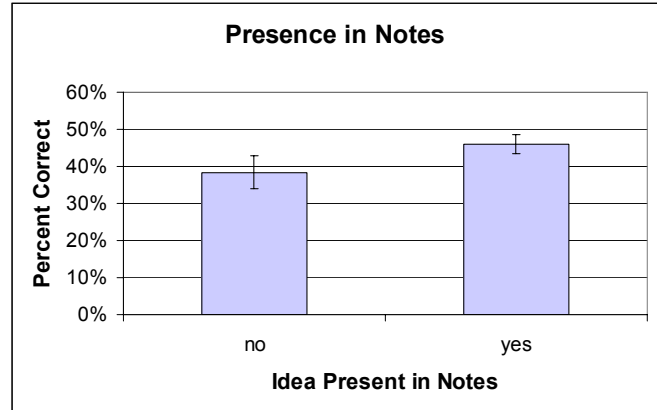


Figure 8: Presence in Notes and Performance on Learning Outcomes

significantly fewer key ideas than the non-novel tools (all $p \leq .01$), but there is no significant difference within the novel or non-novel characterizations. There is a significant overall Key idea wordiness effect [$F(3,58)=9.6, p<.0001$] shown in Figure 7, where Paste is significantly more wordy than all other tools (all $p<.001$). Typing and Restricted are not significantly different, and though Select is significantly less wordy than the two tools ($p<.05$ in both cases), as mentioned above the Select treatment does not have control over wording.

Connecting Notes and Learning

As there were treatment differences in both note-taking behavior and learning, the question arises regarding how these are connected. As each test item was linked to a specific key idea, we could treat each item as a data point linked with a specific key idea, and evaluate note-taking behaviors associated with that key idea by adding them to the full factorial ANCOVA described in the Learning section. Our analyses look for main effects of behavior as well as behavior by treatment, behavior by test-time, and behavior by item-type interactions.

The first analysis suggested by both the data and prior literature regards whether recording an idea influences performance on test items targeting that idea. The novel-tools, which performed worse on learning outcomes than the more traditional tools, also recorded fewer key ideas. Studies cited above indicate that the presence of an idea in notes may in fact influence learning outcomes.

We found a significant effect for presence “in-notes” [$F(1,5362)=3.54, p=.05$], where an item was more likely to be answered correctly if the subject had recorded it in their notes (see Figure 8). This appears to be the case across all tests, as there was no in-note by test-time interaction [$F(2,5362)=.36, p=.69$]. There was a significant item-type by in-note [$F(1,5362)=7.2, p<.01$] interaction, the contrasts of which indicate that in-note was only a significant factor

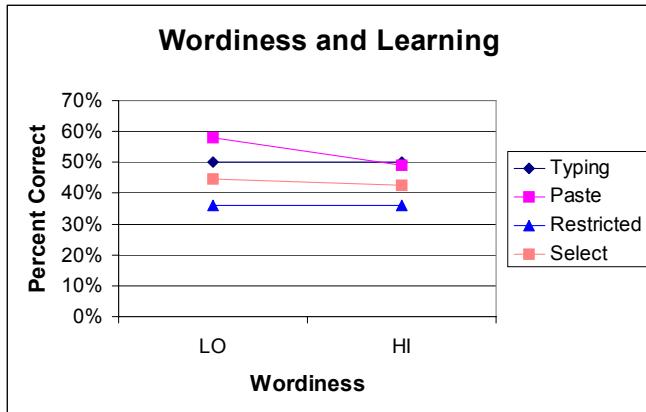


Figure 9: Wordiness and Performance on Learning Outcomes

for multiple-choice items. There was a marginally significant treatment by in-note interaction [$F(3,5362)=2.08, p=.1$], explorations of which showed that the in-note effect was in the same direction for all treatments but paste (for which there was no difference), but only significant for Restricted ($p<.001$).

The treatments were also significantly different with regards to the wordiness with which they recorded key ideas. We explored this phenomenon further, as wordiness has been cited as important in prior literature. In addition, the Restricted tool was designed specifically to reduce the wordiness of copy-paste, with the goal of improving learning. We performed a within-treatment median split for each key idea, which involved grouping notes related to the same key idea for each treatment, and placing each into a “lo” or “hi” wordiness category for that treatment.

We did not find a main effect for wordiness [$F(1,4759)=1.16, p=.28$]. However, there was a marginally significant Wordiness by Treatment interaction [$F(3,4759)=1.99, p=.11$]. Explorations of this interaction show that a wordiness contrast was present only for the Paste treatment (see Figure 9), with more wordy ideas being associated with significantly poorer performance ($p=.01$, for all other treatment contrasts $p>.5$).

Survey Results

The survey included several measures of students’ reactions to the tools. Results of four of these measures most closely related to their satisfaction with the tool are described in Table 1. These questions asked whether students would use the tool in an actual online class, whether the notes they created using their tool helped them study, whether they thought the tool helped them pay attention, and whether the tool promoted learning.

We developed an aggregate “satisfaction” score from these questions. This was computed by giving each subject one point for a positive reaction (i.e. “would use”, “helped learning”) to a question, one negative point for a negative reaction (“wouldn’t use”, “hurt learning”), and no points for a neutral reaction, meaning perfect satisfaction would be a score of four. An ANOVA conducted with treatment as the independent variable and satisfaction as the outcome was significant [$F(3,3)=2.7, p=.05$]. The Selection tool showed significantly lower satisfaction ratings with both Typing and Restricted (both $p<.05$), and marginally lower satisfaction ratings than Paste ($p=.12$).

Students were also asked what they most liked and disliked about the tool they had used. Two-thirds of students who could paste cited it as one of their favorite features. More than half of the people in the Restricted treatment reported disliking the restrictions. Students were also asked what feature they missed the most. Two-thirds of students who could not type reported typing as their most missed feature. More than half of the people who could only type missed pasting. All but 5 students said that paste functionality would be required in any online note-taking application they would use; 4 of these students stated that would never use a note-taking application regardless of its functionality.

DISCUSSION

This experiment was designed to evaluate two hypotheses regarding the impact of selection-based note-taking on learning

We believed that restricting the amount of text a student could select in one copy-paste action would increase their attention to detail, resulting in improved learning gains. Though our Restricted treatment did see the predicted decrease in wordiness relative to the Paste treatment, it also

Treatment	Would They Use Tool?			Did Notes Help Study?		Did Tool Help Pay Attention?			Did Tool Help Learn?			Satisfaction Out of 4
	Yes	No	Maybe	Yes	No	Yes	Neutral	Hurt	Yes	Neutral	Hurt	
Typing	39%	56%	6%	83%	17%	50%	44%	6%	63%	38%	0%	1.4
Paste	41%	59%	0%	76%	24%	33%	33%	33%	41%	41%	18%	0.6
Restricted	44%	50%	6%	83%	17%	33%	33%	33%	61%	22%	17%	1.05
Select	18%	82%	0%	53%	47%	25%	44%	31%	50%	19%	31%	-0.55

Table 1: Satisfaction Survey Results

recorded fewer key ideas and performed more poorly on learning outcomes. In addition, students told us that they disliked the restrictions the interface placed on them. The decreased time on task may be an indicator of this dislike.

Our Selection tool was designed to evaluate whether it was important that students be involved in the creation of their notes or that they view additional representations of learning material. This tool also showed a decrease in the recording of key ideas, and poor performance on learning outcomes. In addition, this was the tool students liked least.

Students appear to have strong opinions regarding how they take notes, which may get in the way of effective note-taking interventions both for research and practical purposes. For example, this study indicates that the most efficient note-taking tool would be a paste-only tool. However, students given this tool want the ability to type. Unfortunately, our previous study suggests that combining typing and copy-paste will eliminate the efficiency gains observed in the paste-only tool.

This study provides no evidence that the Selection tool would be beneficial. As described below, there is evidence that restricting selection may be beneficial. However, the performance of the Restricted tool in this study suggests we use extreme caution, and conduct thorough usability evaluations of any restrictions we impose.

Note-Taking and Learning

Our results provide further evidence that the features included in note-taking applications can influence note-taking behavior. Concretely, the unrestricted ability to paste increases the amount of notes taken with regards to words, ideas, and the wordiness with which ideas are expressed in notes when compared with both typing and more restricted methods of note-taking. Restrictions on selection can significantly reduce the same measures, though not necessarily in beneficial ways.

Our findings do suggest that note-taking tools can impact learning, as the novel tools showed decreased learning gains relative to the tools using standard interaction techniques. Though it may not be terribly surprising to some that tools students dislike inhibit learning, it is important to understand that the tools we design can have an impact on learning outcomes. However, it is unclear from this experiment whether note-taking technology can actively encourage learning, or whether they simply can inhibit learning. This question would have to be resolved with the inclusion in future experiments of treatments in which students do not take notes. Still, the results do give us hope that by discovering optimal note-taking strategies and interaction techniques, we can develop tools that facilitate learning gains.

It also appears that the presence of an idea in students' notes is associated with the likelihood that they will answer

questions targeting that item correctly. Recorded ideas were much more likely to be answered correctly. This result falls in line with previous literature linking the presence of an idea in notes with learning outcomes, and suggests the possibility of modeling students' knowledge by evaluating the contents of their notes.

Note-Taking Interaction Techniques

While the use of the novel tools led to decreased learning outcomes and lower learning efficiency, the more standard tools were similar in their learning outcomes if not in learning efficiency. Interestingly, the standard tools rely on entirely different interaction techniques; keyboard-based typing and mouse-based selection.

Our results are of interest to other note-taking or annotation devices which also use selection as the primary interaction, such as digital highlighting. We have shown here that selection-only techniques can make note-taking more efficient. When unrestricted they do not appear to show decreased learning, and show an increased coverage of total ideas. Highlighting may show similar benefits. However, the paste method has the potential added benefit that the highlighted text is extracted and stored in a (ideally) concise summary that can be used for restudy.

Our results suggest one potential area for improving selection-based learning techniques, while at the same time displaying the difficulty of doing so. When students make larger selections, creating "wordier ideas", they appear to perform worse on learning outcomes. It may be useful to encourage them to make smaller selections. However, the restricted-paste tool described here was shown to be suboptimal.

It is clear that more attention must be paid to the design of the intervention. There is a tradeoff between enforcing behaviors that promote learning and creating a tool that students enjoy using. Therefore a more in depth exploration of potential solutions is warranted. The goal of our restriction was to make students process what they were recording in greater detail by enforcing smaller selections. By deselecting inappropriate selections, we believed students would be required to pay attention to what they were selecting. Though effective with regards to behavior, students used the tool less than standard tools. Multiple designs should be put in front of users, and evaluated with regards to both behavior and desirability. We should also compare restrictions that are enforced with restrictions that are recommended. Simply alerting students to inappropriate selections may be sufficient.

It is also important to note that the increased number of ideas recorded by paste-only users may be detrimental in the long run. This study involved only one module in the context of a 15 module course. Over time, the quantity of notes produced by unrestricted selection may overwhelm students. In fact, this worry has caused most studies of note-

taking while reading to restrict the number of notes their participants can record [e.g. 2, 3]. Identifying ways to make selection more efficient with regards to the key idea to total idea ratio may be useful.

CONCLUSIONS AND FUTURE WORK

This study did not show learning gains with novel note-taking applications. However, the results suggest that it may be possible to develop note-taking tools that actively encourage efficient learning. In fact, the significant learning losses of the novel tools relative to traditional techniques provide evidence that it is possible to influence learning gains through the features included in a standard note-taking applications. In addition, we have shown that selection-based techniques can make note-taking more efficient without associated learning losses. We believe that further studies will allow us to create tools that increase learning gains while at the same time making note-taking more efficient.

Our findings also provide evidence that analyzing students' notes, with regards to the presence of an idea and the wordiness with which it is recorded, can provide insight into the probability that they understand a concept. This information can be used in a variety of ways, such as providing the students with further instruction or restricting how they take notes.

Future work includes evaluations of our note-taking applications in semester-long online courses. It will be interesting to see if behavior changes outside the laboratory setting, when students can return to their notes multiple times and review. We will also conduct further investigations into appropriate restrictions on selections. These will initially require usability studies to identify what type of restrictions, if any, are likely to be adopted by users. Finally, we plan to investigate whether our results apply to online note-taking tools that allow highlighting.

ACKNOWLEDGEMENTS

The authors would like to thank Bill Jerome and John Rinderle for their technical assistance. We would also like to thank Thi Avrahami for her helpful comments on early drafts of this paper. This work has been funded by the Open Learning Initiative and the Pittsburgh Science of Learning Center.

REFERENCES

1. Bauer, A.B., & Koedinger, K.R. Pasting and Encoding: Note-Taking in Online Courses. *Proceedings ICALT*, IEEE Press (2006), 789-793
2. Bretzing, B. H., & Kulhavy, R. W. Note-taking and depth of processing. *Contemporary Educational Psychology* 4, (1979) 145-153.
3. Bretzing, B.H., Kulhavy, R.W. Note-taking and Passage Style, *Journal of Educational Psychology* 73, 2 (1981) 242-250
4. Card, S.K., Moran, T., and Newell, A. The psychology of human-computer interaction *Lawrence Erlbaum Associates*, Hillsdale, NJ, 1983
5. Fisher, J.L., Harris, M.B., Effect of Note-taking and Review on Recall. *Journal of Educational Psychology* 65, 2 (1973), 321-325
6. Howe, M.J. Using students' notes to examine the role of the individual learner in acquiring meaningful subject matter. *Journal of Educational Research* 64, 2 (1970), 61-63
7. Igo, B.L., Bruning, R., McCrudden, M.T. Exploring Differences in Students Copy-and-Paste Decision Making and Processing: A Mixed Methods Study. *Journal of Educational Psychology* 97, 1 (2005), 103-116
8. Kam, M., Wang, J., Iles, A., Tse, E., Chiu, J., Glaser, D., Tarshish, O., and Canny, J.F. Livenotes: a system for cooperative and augmented note-taking in lectures. *Proc. CHI ACM Press* (2005), 531-540
9. Kiewra, K.A., DuBois, N.F. Note-taking Functions and Techniques. *Journal of Educational Psychology* 83, 2 (1991) 240-245
10. Lakmazaheri, S. (2004). SideNote: A Web Annotation Tool for E-Learning in Graphics-Rich Disciplines. *World Conference on E-Learning in Corp., Govt., Health., & Higher Ed.* AACE Press (2004), 2737-2742
11. Marshall, C. Towards an Ecology of Hypertext Navigation. *Hypertext 1998*, ACM Press (1998), Pittsburgh, PA. 40-49
12. Marshall, C.C. & Bernheim Brush, A.J. Exploring the Relationship between Personal and Public Annotations. *Proc DL 2004*, ACM Press (2004), 349-357
13. Nye, P., Crooks, T., Powley, M., & Tripp, G. Student note-taking related to university examination performance. *Higher Education* 13, (1984). 85-97.
14. Palmatier, R.A., Bennet, J.M. Note-taking habits of college students. *Journal of Reading* 18, (1974) 215-218
15. Truong, K. N., Abowd, G.D., and Brotherton, J.A. Personalizing the Capture of Public Experiences. *Proc. UIST'99*, ACM Press (1999), 121-130
16. Van Meter, P.M., Yokoi, L., Pressley, M. College Students' Theory of Note-taking Derived From Their Perceptions of Note-taking. *Journal of Educational Psychology* 86, 3 (1994) 323-338
17. Williams, R.L., Eggert, A.C. Note-taking in college classes: Student patterns and instructional strategies. *The Journal of General Education* 51, 3 (2002), 173-199