Note-Taking, Selecting, and Choice: Designing Interfaces That Encourage Smaller Selections

Aaron Bauer, Kenneth R. Koedinger
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213
1- 412-268-9095
{abauer, koedinger}@cmu.edu

ABSTRACT
Our research develops note-taking applications for educational environments. Previous studies found that while copy-pasting notes can be more efficient than typing, for some users it reduces attention and learning. This paper presents two studies aimed at designing and evaluating interfaces that encourage focusing. While we were able to produce interfaces that increased desirable behaviors and improved satisfaction, the new interfaces did not improve learning. We suggest design recommendations derived from these studies, and describe a “selecting-to-read” behavior we encountered, which has implications for the design of reading and note-taking applications.

Categories and Subject Descriptors
H.5.4 Hypertext/Hypermedia: Miscellaneous. K.3 Computers and Education

General Terms
Design, Experimentation, Human Factors

Keywords
Annotation, Note-Taking, Education, Copy-Paste, Design

1. INTRODUCTION
When creating digital devices for reading, developers can both support traditional paper-based practices and offer novel interactions. The latter is especially true with regards to annotation and note-taking applications. As demonstrated by the XLibris project [22], annotations such as marginal-notes or highlights, which when created are anchored to the document, can be easily extracted and duplicated when created in digital form. For example, a user can highlight material, and then later easily retrieve the highlighted text in a separate document.

A number of studies have focused on understanding how people annotate documents, and the implications for the design of digital support for note-taking and annotation. Digital Library research has looked at ecologies of textbook annotation [17] and formal models of digital annotation [1]. Other studies compare shared and private annotations [17], and looked at how existing annotations affect new readers of documents [25]. Our own research focuses on how digital support for the note-taking process influences learning and behavior in educational settings.

JCDL '08, June 16–20, 2008, Pittsburgh, Pennsylvania, USA.

1.1 Digital Note-taking and Education
Digital text is present in a variety of educational materials, both with regards to entire courses [15] and digital libraries [e.g, 6]. Supporting note-taking and annotation is critical in these environments, as these behaviors have been shown to promote learning. Meta-analyses of a large body of research indicate that both the process of recording a note [13] and reviewing notes [14] support learning. While technology can facilitate note-taking, doing so may reduce the amount students learn in an educational setting.

It is as of yet unclear how digital annotation and note-taking interfaces affect learning and memory. It is clear that interfaces can influence how students create notes. For example, a study of shared annotations indicated that students who intended a note to be shared used more formal wording, and the requirement to attach a comment to a specific piece of text reduced the incidence of more general comments [17].

Our previous research, described in more detail below, provides evidence that while copy-paste based note-taking can be more efficient than typing, it can reduce attention and learning for some students. An initial design aimed at increasing attention was disliked by students, used less, and produced poor learning. This pointed out the need to tradeoff effectiveness with user satisfaction in the design process. As note-taking is an elective behavior, if a restrictive interface is disliked, it will not be used, even if the behaviors it encourages are beneficial.

This paper presents a design study aimed at developing copy-paste interfaces that encourage shorter selections, which are linked with increased learning. It then describes an experimental evaluation of the interfaces. While limited in scope, copy-paste note-taking has other selection-based analogs, such as highlighting and underlining, which the basis of a variety of note-taking applications. Our research also focuses on users' behaviors and goals in order to provide results of interest outside this specific context.

1.2 Copy-Pasting and Wordiness

The courses for which we are developing our interfaces are primarily text-based, and contain interactive examples and self-assessment questions. The initial note-taking interfaces are javascript text-editors embedded in the webpage. Students can record notes either by typing or copy-pasting, and can outline, markup, and edit their notes once they have been created [see Figure 1].

Our first evaluation of this interface was aimed at understanding how using a text-editor to take notes differed from using pencil-and-paper. We found that students took advantage of copy-paste functionality to produce far more notes, which were dominated by verbatim wording [2]. Our second study added learning measures to this comparison, and found that students given the ability to copy-paste forgot more at a week's delay than did students who used pencil-and-paper or who could only type in the text-editor. They recorded ideas in a more wordy fashion, and this wordiness was associated with forgetting, whereas wordiness was beneficial for typing [3].

The effect of wordiness on learning outcomes is not fully understood. An early study found that in lecture students who recorded less wordy notes performed better on tests [10]. However, lecture is time-limited, and recording more wordy notes may result in students not hearing or processing a subsequent idea. Some researchers have used structured matrix note-taking to try to increase wordiness. These structured note-taking applications have only been found to increase note quantity in lecture, not for reading [12]. One study actually found that a digital matrix device that restricted the amount students could enter in any cell performed better than an unrestricted tool [11]. However, it is unclear how results from these structured devices transfer to more general freeform applications, where students are not given the topics that are important for note-taking.

1.3 Wordiness, Attention, and Adoption

Different interfaces have different attentional requirements when it comes to wordiness. As typing requires more time to record additional words, wordiness reflects a greater importance students place on the recorded idea. Copy-pasting more words does not have a similarly increased cost. In this case students may create wordier notes because they are not identifying critical sentences or ideas. In fact, they may not even read the material they are recording thoroughly, choosing instead to select an entire paragraph that seems important. This is reflected in the learning results described above: when it is costly, wordiness results in increased learning, whereas when it is not costly, students forget more.

In order to reduce the cost-free wordiness produced by copy-paste actions, we designed a copy-paste interface that restricted the amount of text students could select in one copy-paste action. Students were limited to ninety-percent of the words in a sentence, and if they went over this limit, their selection disappeared. We were motivated in part by the matrix study mentioned above, which actually improved learning by restricting copy-pasting [11].

Our experimental evaluation of this interface found that it in fact reduced learning. Students disliked the restrictions, and appeared to give up on the tool, recording fewer ideas than they did when typing or using an unrestricted copy-paste tool. Interestingly

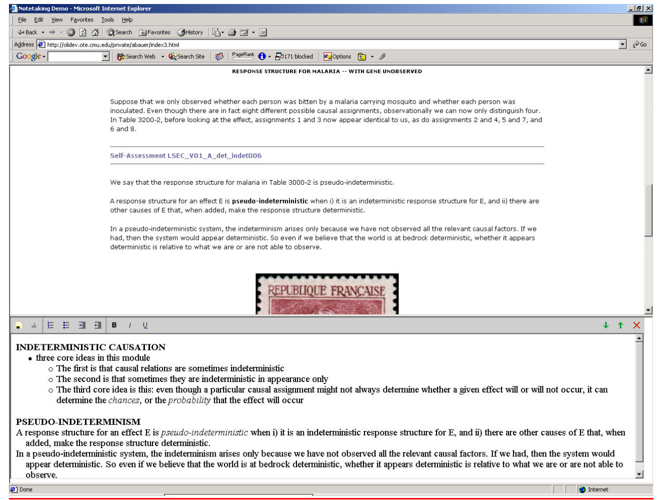

Figure 1: The note-taking interface is a limited functionality text editor embedded in the bottom third of the browser. It allows to outline {bullet, indent, outdent} and markup {bold, italic, underline}. In this study, students could only add organizational characters (such as dashes or parenthesis) to their notes.

students using the tool that only allowed them to copy-paste, but did not restrict selection-size, were more efficient than typists, learning the same amount in less time. However, when they still performed worse on learning outcomes for ideas they recorded in a more wordy fashion [4].

1.4 Elective Technology

While copy-pasting appears to be more efficient than typing, the wordiness effect demonstrates the potential for improving copy-paste note-taking even further. If we can encourage copy-pasters to focus on what they are recording, we may be able to create a tool that is more efficient and produces better, rather than equivalent, learning. However, our previous results indicate the difficulty of designing such a tool. Students did not like the interface, and chose to use it less. It is clear that if students are to be encouraged to record notes in beneficial ways, a restrictive interface must be designed very carefully. At issue is the elective nature of note-taking. In the matrix study [11], note-taking was required, whereas here students could choose not to take notes if they did not like the interface. Though we could require note-taking as well, in most educational settings note-taking is not enforced, and it is required even less outside of educational contexts.

Adoption is not a problem unique to note-taking. Optional comprehension checks in online courses are another example. Though positively associated with learning outcomes, they are not always fully utilized by students [21]. Computers can be used to increase the use of such optional behaviors. Hausmann and colleagues showed that self-explanation, an optional behavior in which students generate inferences and personal understanding of material with which they are presented, can be prompted by software, and that such prompting increases learning [9].

While studying whether presenting multiple documents for reference simultaneously improved essay writing, Wiley found that, given the option, students chose to view one at a time. In

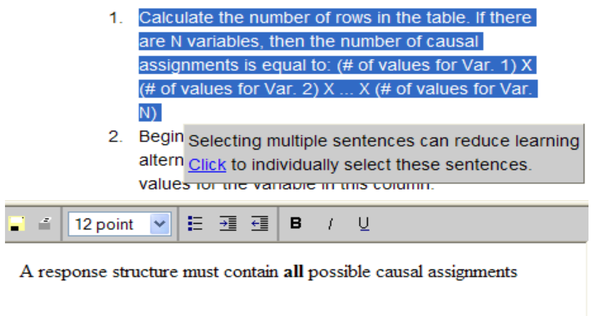

Figure 2: Example of a selection-based interface recommending that the user select fewer sentences.

order to experimentally evaluate the value of simultaneous presentation, the interface had to be obligatory [23]. While the intervention produced positive results, it is clear that if left optional, it would remain unused, and students would not benefit. It is unclear whether any design concessions one might have to make to encourage simultaneous document viewing would retain the positive results found with the obligatory interface.

It is also important to note that user satisfaction will be an especially important factor in digital library applications not developed specifically for educational purposes. In these cases where learning or memory might not be primary goals, the tradeoff between learning and satisfaction is weighted almost entirely in the direction of satisfaction.

2. DESIGN STUDY

In order to develop a restrictive interface more likely to be adopted than the one reported in [4], we must tradeoff effectiveness with user satisfaction. We conducted a rapid design study in order to accomplish this goal. The study was intended to explore the impact of different interface elements and study how students interacted with note-taking applications. The best interfaces, which showed highest user satisfaction while still reducing selection-size, would be included in a rigorous experimental study we will describe in the next section. The first concession we made was to base all restrictions on the sentence level, whereas the previous design limited students to a percentage of a sentence. This was done because as sentences are easily recognizable units, restrictions based on them would be less confusing than restrictions based on a percentage of words in a sentence. The first stage of our study included the old interface for comparison purposes in order to confirm this decision.

2.1 Method

We believe that iteration and observation are crucial to both understanding user resistance to intervention and designing interfaces with a greater likelihood of adoption. We followed a process whose first step involved the identification of a set of interface elements and intervention styles that can be manipulated. Interfaces were created that addressed these variables, and were then user-tested. The user-tests were analyzed with respect to the variables manipulated, and candidate interfaces were created. These candidates were then tested with users. Such iteration is a basic and effective principle of interface design [19]. Our user-tests utilized the think-aloud protocol, which asks users to express their thinking verbally, and is a standard observational technique for usability derived from cognitive science [5].

Think aloud, iteration, and observation are not novel in the educational community either. In fact, both form a large part of the recent work on “design studies” [7]. However, these design studies are large scale in classroom interventions, whereas this paper is focused on designing optional interfaces for experimentation.

We used the think-aloud protocol to conduct our user-tests. Each participant completed a 15-page module in Causal and Statistical Reasoning using three different interfaces to record their notes. As motivation for note-taking participants were given a short 9-question quiz during which they could review their notes. Students were then interviewed regarding their note-taking behavior. Screen and audio captures were recorded for everything but the quiz.

Six subjects participated in first phase user-tests, and 5 participated in the second phase. All 11 were recruited using a university bulletin board. Participants included undergraduates, graduate students, and staff members at a major university. Participants were tested using Internet Explorer on a Windows XP desktop with a 17-inch monitor. In our descriptions of the data, we identify relevant subjects by using their subject ID (i.e. Subject 1 = S1).

2.2 Initial Designs: Exploring Variables

During the first phase we identified four main variables for evaluation. The first involved intervention style, which contrasts *restriction* and *recommendation* distinction. While a restriction does not allow a student to make a lengthy selection, a recommendation informs a student when their selection may be too lengthy. Though a restriction guarantees compliance, it may suffer with regards to adoption when it does not allow users to complete desired tasks.

The next variable we identified regards *in-process vs. post-process* intervention. A user can either be informed about an inappropriate selection when it occurs, or after the user has finished creating a selection of inappropriate length.

The third variable regards how to treat inappropriate selections when they are created in restrictive interfaces. The interface could either automatically reselect a single sentence, or individually select every sentence within the user’s selection, allowing the user to copy-paste only one at a time. While the former would require an arbitrary decision regarding which sentence to select, the latter would not require identification of a single important sentence on the part of the user.

The final variable regards how to provide feedback. The changes in selection themselves serve as an implicit type of feedback. The interface could also popup a warning when students create inappropriate feedback, either when it happens or after the selection has been created.

In the first phase we created seven interfaces using different combinations of the above variables.

Hard-Stop: Once students started a selection, they could not select beyond the boundaries of the initial sentence. Initially the interface popped up a message when students reached a boundary and did not allow them to select any further. The popup was abandoned after initial user complaints.

Reselect-Sentence: After a selection was completed, if the selection contained multiple sentences, the first sentence was automatically reselected.

Reselect-Multiple: Same as above, but all sentences in the selection were reselected individually, so the student would have to copy (but not select) each one individually [figure 2].

Recommend-Sentence: Once the sentence boundary was crossed, the interface popped up a recommendation that the student reselect. It gave the option of clicking a link to have the interface automatically reselect the first sentence.

Recommend-Multiple: Same as above, but the link reselected all sentences.

Click-Select: This explored the possibility of disguising the restriction as a feature. Students could select a sentence by clicking on it, but could not create any other selections.

Finally, the restricted copy-paste interface from our previous experiment was included alongside a tool allowing unrestricted selection. The former interface allowed students to select no more than 90% of the words in a sentence. If an inappropriate selection was made, it simply disappeared.

2.3 User-Tests of Initial Interfaces

One of the more unexpected results from our initial user-tests was the discovery of both unintentional and intentional selection behaviors that had nothing to do with note-taking. At least 3 subjects accidentally selected multiple sentences when their mouse deviated slightly from the line they were intending to select. More interestingly, 6 of our 11 participants selected text in order to facilitate reading. These selections ranged from individual words or phrases that were emphasized in speech to selections spanning multiple sentences. During the interviews, participants stated such behavior served to help them concentrate (S11) or reminded them what they had left to read (S6). These findings played a large part in some of the conclusions derived from these initial user-tests, which we will now describe.

It is preferable to give explicit feedback post-process than in-process. In-process feedback can be triggered by the unintentional selection errors described above. In addition, the feedback severely disrupted the note-taking process of the first user of the hard-stop restriction, who expressed quite a bit of frustration ("I know, I didn't mean to…You're killing me!"). This led us to eliminate the popup for subsequent users of that interface, as their inability to select more text already served as implicit feedback. At the very least, popups should not be given as feedback in-process, as they could be prompted by accidental selections or selections for non-note-taking purposes. Even post-process feedback serves no purpose when selections are being made for reading purposes alone.

Changing the user's selection can be inappropriate. When first using the tool that reselected a single sentence, S3 stated "this is sick!" During his interview he stated a preference for the hard-stop restriction, as the after-single one "is allowing me to select something, and then it's saying no! It's like giving me something and then taking it back." When subjects do not realize a reselection has occurred, it can result in *transcription errors* as they believe the selection they intended to make was pasted. Often students will not read what they have copied in their notes, so the transcription error becomes permanent (S4, S6), and their notes do not reflect their intentions.

	Design Guidelines
1	Give explicit feedback after, not during, the creation of a selection.
2	Do not modify selections once they are created.
3	Where possible, present restrictions as features
4	Feedback should not offer reselection functionality

Table 1: Design guidelines derived from user-testing.

Changing the user's selection also appears to affect selecting to read behavior. S3, who was displeased with the single sentence reselection, was one of the most frequent users of selection to read. Interestingly, he was able to use the hard-stop restriction to accommodate his behavior.

Our data does not indicate whether recommendations would effectively discourage multiple-sentence selections. Of the four users who tested a recommendation tool in the first phase, only one clicked once in either of the recommendation tools. During the interview one user (S6) stated a dislike for the recommendations, saying they "served no purpose", and would not be used. While it may be that feedback serves as a constant reminder, it is clear that such feedback should not be given for non note-taking behaviors.

The click-sentence was promising. It was the favorite of two subjects, and no student who used it like it the least. S3, who did not use the click-select interface to record notes, discovered that he could select a sentence using the hard-stop tool by double-clicking on it, and subsequently used this feature extensively. *Implementing restrictions as novel features appears to be a useful approach.*

The first two users demonstrated the deficiencies of our old interface, which eliminated selections that contained more than 90% of the words in a sentence. This somewhat arbitrary restriction turned out to be quite unintuitive. While subjects attempted to comply with the restriction by selecting within a sentence, they would still select too much text, and their selections disappeared (S1, S2). Both users given this tool expressed frustration with it, and one user (S2) reported giving up on note-taking. Procedural workarounds, such as selecting the entire sentence in two parts, were developed.

2.4 Second Design: Candidate Interfaces

We developed three candidate interfaces for evaluation during the second phase. All interfaces followed the guideline of not giving explicit in-process feedback. A revised c*lick-select* interface combined the features of the click-sentence and hard-stop interfaces from the first iteration. The initial click-sentence was well regarded even by people who demonstrated select-to-read behavior. However, it did not allow selecting small pieces of text. This was important both to support select-to-read and because the goal of this design is to promote smaller selections. It also satisfied the principle of not modifying a selection after its creation. This very combination of features was suggested by S6 after using the hard-stop tool while studying and being shown click-select during the interview.

The second interface was a modified version of the Reselect-Multiple tool. While this violated the guideline against changing the user's selection, all of the text the user originally selected would continue to be selected. We believed this would reduce frustration, and provide support for multiple sentence selecting to read. We hoped to reduce transcription errors by clearing the clipboard when sentences were reselected. This meant that if the student failed to copy the selection, no notes would be placed in the notepad if they pasted. We hoped users would be more likely to notice nothing being pasted than they were to notice an incorrect sentence being pasted. We also hoped to reduce transcription errors by highlighting the individual sentences on mouse over so that students would be more likely to notice what they are and are not copying.

The final interface was a revised recommendation tool. In this tool warnings were only displayed upon completion of a selection, in order to avoid warnings caused by selection errors. The interface only starts to display warnings when users copy or start to drag an inappropriate selection. This avoids giving warning for behaviors such as selection-to-read. However, if students do copy or drag multiple sentences into their notes it then "nags" students whenever they select multiple-sentences. If they begin to copy single sentences again, the interface no longer displays the nag popup. This is done by comparing the number of multiple sentence copy-paste actions with the number of single sentence copy-paste actions. The nag popup only appears when there are more multiple-paste actions than single-paste actions. We were hopeful that this carrot-and-stick approach would encourage shorter selections.

2.5 User-Tests of Candidate Interfaces

The click-and-select interface was the preferred of the two restricted tools. It was used for selecting both entire sentences and parts of sentences for note-taking (S7, S8, S9, S11). It also supported selection-to-read behaviors for several students (S7, S11). The multiple-select tool continued to produce transcription errors. When S7 forgot to copy the last sentence in a 3-sentence selection, she pasted the second selection twice. While she caught this transcription error, the evidence above indicates not all students would.

The recommendation tool showed some promise, though. *the option to reselect text should be abandoned, as it is not utilized*. Several subjects whose behavior had caused the nag screen to appear made it disappear by pasting individual sentences, though it is unclear from the behavioral data that this had anything to do with the warnings. During the interview one (S8) stated that they selected individual sentences in order to avoid the nag popup. When asked why the recommendation tool was her favorite, S7 stated "I like the warning that copying too much was bad, because then you can wind up copying things that are just really framing and not the essential." However, only an experimental evaluation can determine whether it affects behavior by making students aware of potentially negative behaviors.

3. EXPERIMENTAL EVALUATION

We compared the best restriction and recommendation interfaces in our experimental evaluation. The click-select tool was the preferred restricted tool. The recommend tool was included with two modifications. First, the link to automatically reselect text was abandoned, as it was not used. Secondly, rather than basing the nag popup on a count, it went away as soon as the user copy-pasted an appropriate selection. These two tools were compared with a condition in which students took notes using an unrestricted copy-paste interface, and a condition in which they were not allowed to take notes.

In a slight variation to our previous studies, this study quantifies wordiness with regards to multiple-sentence vs. single-sentence selections, as our tools were designed to promote single-sentences.

3.1 Hypotheses and Goals

This experiment was designed to evaluate several hypotheses.

H1: The novel interfaces will produce fewer multiple-sentence selections than the unrestricted copy-paste tool. The reduced coverage of key ideas seen in the previous experiment will not occur in this experiment.

H2: The recommendation tool will see higher satisfaction ratings than the click-select tool because it did not restrict users, but will result in more multiple-sentence selections, as not all users will comply with the recommendations.

H3: The novel interfaces will not see the negative user satisfaction ratings observed in our previous experiments.

H4: By increasing the attention students pay to the note-taking process, the novel interfaces will result in increased performance on learning outcomes.

We were also interested in using a larger subject pool to address findings from our design study. Of primary interest was determining the frequency of "selection-to-read" behaviors. This would be viewed both through behavior, and by asking about the behavior in the post-test survey.

3.2 Design and Subjects

The study presented here compares students taking notes in an online classroom using four treatments: "Click-Select", "Recommend", "Unrestricted Copy-Paste", and "No Notes". All interfaces are described above. The study follows 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.

In order to avoid a potential self-selection bias present in our early studies, where students in the Paste treatment could type when pasting would be most detrimental to learning, no interfaces included the ability to type.

A total of 53 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; their data was not included in the analyses described here. No students reported being familiar with the course materials. Participants were paid per hour participated.

3.3 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.

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

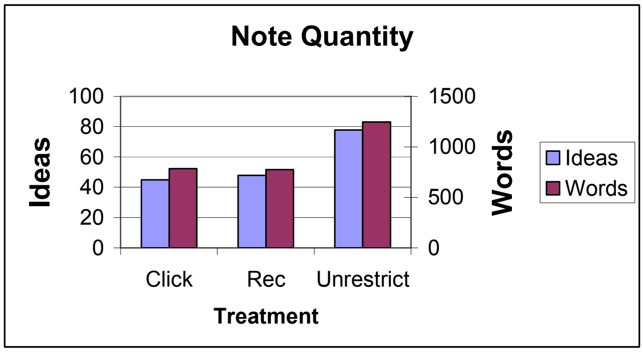

Figure 4: Note Quantity. The unrestricted condition recorded significantly more notes, with respect to both words and ideas, than the other two note-taking conditions.

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. Students in all treatments could copy-paste within the notepad, and type dashes, parentheses and similar outlining characters.

Each quiz contained 21 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 9 multiple-choice items and 12 free response items.

3.4 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. 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

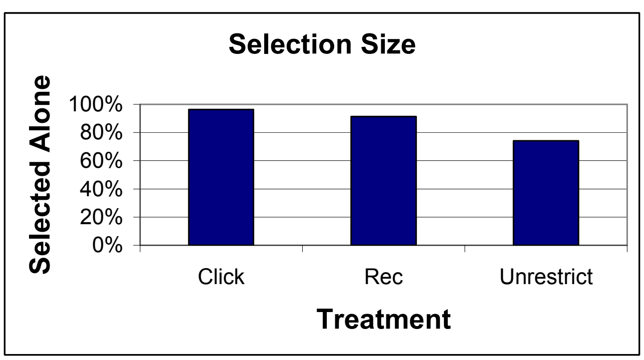

Figure 5: This graph shows the percentage of ideas that were present by themselves in notes. The unrestricted condition had significantly fewer single-sentence selections.

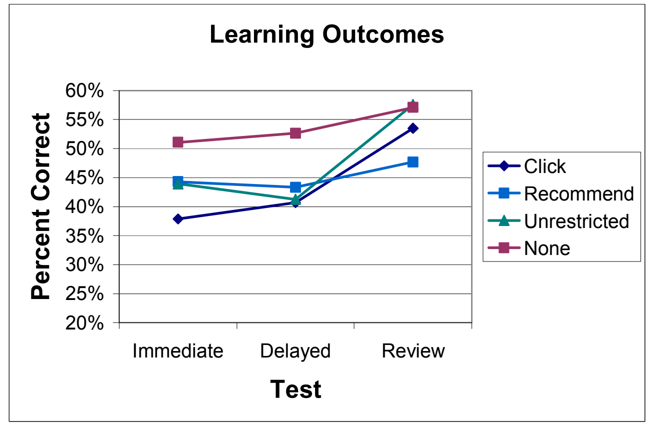

Figure 6: Overall Learning Results. There were no significant differences on individual tests. The click and unrestricted tool were the only interfaces to benefit significantly from review.

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. 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. Students in the no-notes condition were asked to mentally review, which is a standard procedure in note-taking studies. They were then given the final test. After the test they were given a short survey regarding their experience.

3.5 Dependent Measures

We collected several dependent measures, including completion time, measures of note quantity, logs of selections and copy-paste actions, and 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. Notes were parsed and coded in Microsoft Excel by means of a VBA script written for our previous experiment.

3.6 Results

3.6.1 Behavior

ANOVAs were conducted on each note-taking measure, with as the only independent measure. Pretest was not included in the model, as it was not found to be a significant covariate. Time on task was not found to differ by treatment [$F(3, 49)=.211$, $p>.8$], so it cannot be responsible for any observed differences.

There were significant effects observed for the total number of ideas [$F(2,37)=3.19$, $p=.05$] and the total number of words [$F(2,37)=3.16$, $p=.05$] recorded by each tool [figure 4]. In both cases, contrasts showed that the unrestricted tool recorded a greater quantity than the other tools, which were not significantly

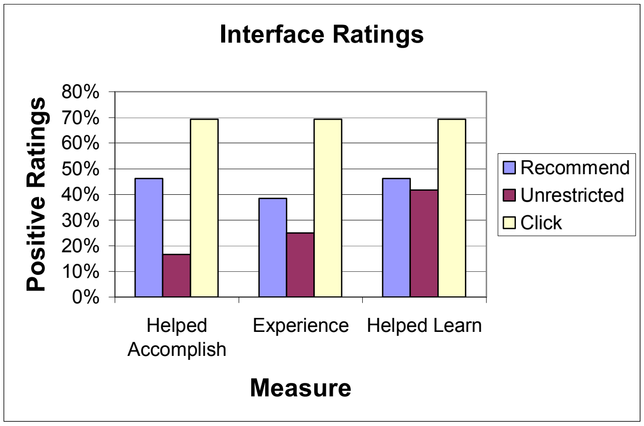

Figure 7: Interface ratings. Click-select was significantly higher for accomplishing goals and experience. When averaged together, the ratings are also significantly different.

different from each other. There was not a significant difference with regards to the wordiness with which each tool recorded its ideas [F(2,37)=2.07, p=.14]. However, there was a significant difference in whether the tools copy-pasted individual ideas or groups of ideas in one selection [F(2,37)=8.17,p=.001]. Once again, contrasts indicate the unrestricted tool was significantly more likely to record ideas together, rather than selecting each one individually [figure 5].

3.6.2 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 (No notes vs. recommend vs. unrestricted vs. click) and test-time (immediate vs. delay vs. review) as a within-subjects variable. Item-type (Multiple-Choice vs. Free-Response) was not found to be a significant within-subjects variable, and was left out of the analyses described here. There was not a significant effect of treatment [F(3,45)=1.11, p>.3]. There were significant effects for pre-test [F(1,3242)=44.5, p<.001] and test-time [F(1,2342)=12.09, p<.001]. There was a marginal test-time by treatment interaction [F(6,3242)=1.8, p=.09]. This was between the delayed and review tests on the second day, where only the click and unrestricted tool receiving any benefits from review (p<.01 for both within-treatment contrasts).

We also did analyses connecting note-taking behavior with learning. 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 ANCOVA described above. Our analyses look for main effects of behavior as well as behavior by treatment interactions.

Whether an idea was ever alone in notes was marginally significant F(1,2113)=2.85, p=.09, and interacted significantly with treatment [F(2,2113)=6.78, p=.001]. Contrasts for the treatment interaction indicated that being alone was a significant factor for the unrestricted condition and the click condition. Though being alone was positive for the unrestricted condition, it was negative for the click condition. Presence was found to be a significant positive factor [F(1,2347)=33.41, p<.0001], and interacted with condition [F(2,2347)=7.16, p<.001]. Investigating the treatment interaction indicated that presence was only a significant factor for the two novel tools.

3.6.3 Survey: Attitude and Conscious Behavior

On a 7-point Likert scale, students using note-taking tools were asked to rate a) the degree to which the interface they used allowed them to *accomplish* their goals b) their *experience* using the interface from frustrating to pleasant, and c) whether the tool increased or decreased *learning*. For analysis purposes, the rating data for all measures was consolidated into a nominal variable with two values: positive/above neutral (greater than 4 on the Likert scale) or at or below neutral. Standard chi-square tests were performed.

The click select tool appears to be better received than the other interfaces [figure 7]. There was a significant difference for accomplishment, $\chi^2(2, 38)=6.9$, p=.03, a marginal difference for Experience $\chi^2(2, 38)=5.2$, p=.07, and no difference for learning $\chi^2(2, 38)=2.23$, p>.3. Averaging scores across all measures is also significant $\chi^2(2, 38)=9.09$, p=.01.

The survey asked our fifty-one participants questions regarding note-taking behavior. To validate the select-to-read behavior from our design study, we included a question regarding whether students in general used selection for purposes other than reading, and if so what were those purposes. Thirty-nine students reported this behavior. Thirty-two of them said they select text to help them focus while reading, 16 reported selecting text to make it easier to read, and 10 reported using selections as a bookmark while reading.

Analysis of selection logs also indicates that two of the twelve subjects in our no notes condition used selection-to-read extensively (one selecting text 30 times, another 70 times). Another 3 selected text less than 3 times, while the remainder never selected text.

We also asked students why they took notes. Forty-two reported the process of taking notes helped them remember material, and 20 stated this was their primary reason for taking notes. Forty-six students reported taking notes to review them later, 18 of whom said this was their primary purpose.

3.7 Discussion

Our data supported hypothesis 1. Both novel tools made significantly fewer multiple-sentence selections than the unrestricted tool. While this is not surprising with the click-select tool, the recommend tool allowed unrestricted selection, so the recommendations appear to have been effective.

There is no support for hypothesis 2. There was no difference in number of multiple-sentence selections between the two novel tools. It appears the recommendations were effective at dissuading users. Interestingly, the restricted click-select tool even enjoyed higher user satisfaction than the recommend tool.

While students the novel interfaces recorded fewer total ideas than the unrestricted interface, they recorded an equivalent number of key ideas. While there is an association between the presence of a key idea in notes and learning outcomes, there is no association between total note-taking quantity and performance, so focusing on key ideas may be an appropriate behavior. In fact, it may be beneficial over a semester, when a larger quantity of notes may be unmanageable.

The data supported hypothesis 3, as not only did the recommendation tool show equivalent user satisfaction ratings to the unrestricted tool, but the click-select tool showed better user satisfaction than any other tool. Our design study appears to have been effective with regards to user satisfaction as well. In our previous experiments, students disliked our interventions. In this study, students enjoyed the experience of using the click-select tool more than the other interfaces, and thought it was more useful in accomplishing their goals. This is especially surprising considering the unrestricted tool allowed them more freedom. It appears that hiding the restriction in the guise of a feature was an effective design principle with regards to user satisfaction. The recommendation tool did not differ from the unrestricted tool with regards to user satisfaction.

Our data do not support hypothesis 4. There was not an overall learning advantage for note-taking for any condition. In fact, taking notes in this experiment was no better than just reading the material. While reviewing was valuable for the unrestricted and click-select condition, it did not place their performance above students who did not take notes. Thus the overall importance of note-taking for these course materials is questionable. Students, however, believe note-taking was valuable for this experiment. They also believe that both the process of taking notes and having notes for review is valuable.

In addition, students in the unrestricted condition performed better when they recorded ideas individually. However, while our interfaces were effective in encouraging single-sentence selections, this did not improve learning. These results suggest that designing interfaces to encourage note-taking behaviors that are associated with learning may not be effective.

Selection-to-read behaviors were confirmed in this study. We observed some students who were not taking notes at all frequently selecting text. A majority of students report selecting text while reading outside of this experiment. The dominant reason for using selection was to help students focus, though others selected text to facilitate reading poorly designed text, and a group of students report using selected text as bookmarks.

4. GENERAL DISCUSSION

4.1 Recommendations from Design Study

We found the design study to be an inexpensive technique for designing user-friendly interfaces that encourage specific behaviors. The interfaces produced the desired behaviors without suffering the dissatisfaction observed with our previous interface. In fact, the click-select tool was more popular than the unrestricted tool. In addition, the interfaces resulted in the intended behaviors.

The experimental results confirm some of the recommendations from the design study (see Table 1). The success of the click-select tool indicates that introducing a restriction by offering a new feature is a valuable design solution. Feedback should not be given while users are selecting text, as it will distract them. If the interface permits a selection, it should not be modified, as users may not be conscious of these modifications, which can result in transcription errors. Restrictions should instead be imposed in-process. This is especially effective with regards to user-satisfaction if the restrictions are in the guise of a new feature. Giving students the option to automatically reselect an appropriate phrase is not useful, as students will not use that functionality. When done unobtrusively, nagging users can encourage desirable behaviors without reducing user satisfaction. However, our nag interface was still not as satisfying as our restrictive interface.

4.2 Selecting to Read

These studies identified several behaviors of interest and their consequences. When recording notes, students may not read what they have recorded. We also found that selection errors, often caused by small motor errors, often led to transcription problems. Selection errors resulted in permanent note-taking errors, as students never realized they had transcribed the wrong material.

In our design study, we observed students selecting text without having any intention to record it in their notes. We confirmed this behavior with a larger subject pool in the experimental study. Not only did the behavior persist, but also a majority of students give reasons for selecting digital text while reading. They report selecting text to help them focus while reading, selecting text when they find the font difficult to read, and using selections as a temporary bookmark on the page they are reading.

This behavior may not be as abnormal as one might expect. In a study comparing reading digital documents on a tablet with reading a paper, the authors point out “lightweight navigation” features present in paper that are missing in their tablet interface. One of these is the ability to narrow or broaden focus, which readers of magazines accomplish by folding or reorienting the paper. Their tablet readers do not demonstrate similar behaviors, as they are not available in the interface [18]. The selection-to-read behaviors observed in our study seem to accomplish the same goals of focusing attention. Allowing users of reading appliances or interfaces the ability to select text may be one way of supporting narrowing of focus.

4.3 Interventions and Learning

Our previous research linked shorter selections with improved learning outcomes. While the current study still indicates that for an unrestricted tool shorter selections are better, our interventions did not increase learning even though they reduced selection size. It appears that the benefits achieved by shorter selections are not realized when such selections are imposed by the interface rather than chosen by the student. It may then not be preferable to intervene to change note-taking behavior when learning is the goal. Still, the results suggest at the very least that designers consider whether the features they include in their applications encourage longer selection.

However, if we leave note-taking unrestricted, we still have data linking presence in notes and how ideas are selected with learning outcomes. This data can be used to update models of student knowledge or give further instruction. For example, if a student does not select a key idea, or selects it only as part of a larger selection containing multiple selections, that student is less likely to perform well on learning outcomes associated with the key idea. We can use this information to update a model of the student’s knowledge, for example using intelligent cognitive tutors [8]. Alternatively a course could also give the user self-assessment questions targeting that idea, or a library could give additional readings that target ideas students are less likely to know.

5. CONCLUSIONS

In this paper we described a design study intended to develop note-taking interfaces that reduce selection size in copy-paste note-taking in order to improve learning outcomes. The experimental evaluation of the interfaces we produced indicated the interfaces achieved their behavioral goals while maintaining, and even improving, user satisfaction, but did not improve learning as expected. Note-taking in general was not seen to benefit users with regards to learning. These results may be tied somewhat to the learning materials used in this study. Note-taking studies in general have found conflicting results with regards to learning outcomes.

While it did not produce interfaces that increased learning, the design study proved to be a low cost effective method of developing interfaces that encouraged desired behaviors while maintaining user satisfaction. In addition to producing a set of guidelines for user-friendly note-taking interfaces, it identified a selecting to read behavior that would be important for developers of any reading or note-taking application.

While this research has been focused on copy-paste based note-taking, the results may apply to other annotation interfaces. In particular, design considerations should apply to other selection-based techniques such as highlighting or underlining, and learning results may apply as well. Our future work will compare copy-pasting with highlighting.

6. ACKNOWLEDGEMENTS

The authors would like to thank Bill Jerome and John Rinderle for their technical assistance. This work has been funded by the Open Learning Initiative and National Science Foundation grant SBE-0354420 to the Pittsburgh Science of Learning Center.

7. REFERENCES

[1] Agosti, M. and Ferro, N. 2007. A formal model of annotations of digital content. ACM Trans. Inf. Syst. 26, 1 (Nov. 2007), 3.

[2] Bauer, A., and Koedinger, K. “Developing a Note Taking Tool from the Ground Up”. Ed-Media 2005. AACE Press, 4181-4186.

[3] Bauer, A., Koedinger, K.R. Pasting and Encoding: Note-taking in Online Courses. In Proc. IEEE ICALT 2006. 789-793

[4] Bauer, A. and Koedinger, K. R. 2007. Selection-based note-taking applications. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (San Jose, California, USA, April 28 - May 03, 2007). CHI '07. ACM, New York, NY, 981-990.

[5] Boren, T., Ramey, J., Thinking aloud: reconciling theory and practice, IEEE Transactions on Professional Communication 43, 3 (2000) 261-278

[6] Brusilovsky, P., Farzan, R., and Ahn, J. 2005. Comprehensive personalized information access in an educational digital library. In Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries (Denver, CO, USA, June 07 - 11, 2005). JCDL '05. ACM, New York, NY, 9-18.hey

[7] Collins, A., Diana, J., Bielaczyc, K, Design Research: Theoretical and Methodological Issues, The Journal of the Learning Sciences 13, 1 (2004) 15-42

[8] Corbett, A. T., Koedinger, K. R., & Anderson, J. R. (1997). Intelligent tutoring systems (Chapter 37). M. G. Helander, T. K. Landauer, & P. Prabhu, (Eds.) Handbook of Human-Computer Interaction, 2nd edition. Amsterdam, The Netherlands: Elsevier Science.

[9] Hausmann, R. G. M., & Chi, M. T. H.. Can a computer interface support self-explaining? Cognitive Technology, 7(1), (2002) 4-14.

[10] 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

[11] 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

[12] Kiewra, K.A., Dubois, N.F., Christensen, M., Kim, S.I., Lindberg, N., A more equitable account of the note-taking functions in learning from lecture and from text, Instructional Science, 1989, 18, 217-232

[13] Kobayashi, K. 2005. What Limits the Encoding Effect of Note-Taking? A meta-analytic examination., Contemporary Educational Psychology, 30, 242-262

[14] Kobayashi, K. 2006. Combined Effects of Note-Taking/-Reviewing on Learning and the Enhancement Through Interventions: A meta-analytic review. Educational Psychology 26, 3 459-477

[15] Livingston, A., Ed., Wirt, J., (Eds.). The Condition of Education in Brief 2004. National Center for Education Statistics, U.S. Department of Education (2004), Retrieved August 8, 2004 from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2004076

[16] Marshall, C. Towards an Ecology of Hypertext Navigation. Hypertext 1998, ACM Press (1998), Pittsburgh, PA. 40-49

[17] Marshall, C.C. & Bernheim Brush, A.J. Exploring the Relationship between Personal and Public Annotations. Proc DL 2004, ACM Press (2004), 349-357

[18] Marshall, C. C. and Bly, S. 2005. Turning the page on navigation. In Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries (Denver, CO, USA, June 07 - 11, 2005). JCDL '05. ACM, New York, NY, 225-234

[19] Nielsen, J. Iterative User-Interface Design, IEEE Computer 26, 11 (1993) p.32-41

[20] Palmatier, R.A., Bennett, J.M., 1974. Notetaking Habits of College Students, Journal of Reading, 18, 215-218

[21] Scheines, R., Leinhardt, G., Smith, J., and Cho, K. Replacing Lecture with Web-Based Course Materials. Journal of Educational Computing Research, 32, 1, (2005) 1-26.

[22] Schilit, B. N., Price, M. N., and Golovchinsky, G. 1998. Digital library information appliances. In Proceedings of the Third ACM Conference on Digital Libraries (Pittsburgh, Pennsylvania, United States, June 23 - 26, 1998). I. Witten, R. Akscyn, and F. M. Shipman, Eds. DL '98. ACM, New

York, NY, 217-226. DOI= http://doi.acm.org/10.1145/276675.276700

[23] Wiley, J., Supporting Understanding Through Task and Browser Design. Proc. of the Twenty-third Annnual Conference of the Cognitive Science Society, Hillsdale (2001) 1136-114

[24] 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

[25] Wolfe, J. L. 2000. Effects of annotations on student readers and writers. In Proceedings of the Fifth ACM Conference on Digital Libraries (San Antonio, Texas, United States, June 02 - 07, 2000). DL '00. ACM, New York, NY, 19-26.