Designing Note-taking Interfaces for Learning

Aaron Bauer
June 2008
CMU-HCII-08-103

Doctoral Dissertation
Human-Computer Interaction Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA USA
Carnegie Mellon University, School of Computer Science

Thesis Committee:
Kenneth R. Koedinger, chair
Jodi Forlizzi
Chris Neuwirth
Charles Perfetti

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Abstract

Note-taking is a common behavior for students both while reading and while attending lecture. An extensive history of research dating back to the early 20th century has shown that both the process of note-taking and having notes to review promote learning. As technology changes the ways learning materials are delivered, note-taking applications are being built for digital environments. While these applications have been shown to change how students take notes, few studies exist regarding the impact these changes in behavior have on the positive learning gains achieved through traditional note-taking. The research in this thesis addresses this problem by comparing both behavioral and learning outcomes of different selection-based note-taking applications, such as copy-paste and highlighting. It is also designed to offer insight into the relationship between note-taking and learning, with particular attention being paid to theories of focusing and elaboration.

The results of this work indicate that not only does the functionality included in a note-taking interface affect the quality of students' notes, but it also can have an impact on learning. The research provides evidence that one of the potential benefits of technology is increasing the efficiency with which students can take notes. It also finds that students given more efficient interfaces, that allow them to learn the same amount in less time, want features that increase time without benefiting learning. Finally, it points out the issue of lack of adoption of optional interfaces designed to encourage student behavior's associated with learning gains, and describes a design process that addresses this problem.

Acknowledgements

I would like to thank my advisor, Kenneth R. Koedinger, for his support, advice and insight during the past six years. The work reported in this thesis was aided immeasurably his guidance. I consider myself quite fortunate to have had the opportunity to work with him.

The breadth and depth of knowledge provided by my committee was also invaluable. Jodi Forlizzi was a strong influence not only on improvements on the design processes reported here, but in my overall growth in design. The connections Chris Neuwirth helped me make between my work and the reading comprehension literature were particularly helpful in interpreting my results, and her feedback on my user studies assisted in the collection of better data. Charles Perfetti helped ensure that the work stay grounded in educational theory, and provided excellent guidance during the final phase of my thesis work.

Many people have contributed to my work and provided support over the past few years, including: Lisa Anthony, Daniel Avrahami, Thi Avrahami, Ryan Baker, Jo Bodnar, Matthew Easterday, James Fogarty, Darren Gergle, John Graham, Dave Holstius, Gary Hsieh, Bill Jerome, John Kembel, Andy Ko, Queenie Kravitz, John Rinderle, Ido Roll, Peter Scupelli, Cristen Torrey, Jake Wobbrock, Ruth Wylie and the rest of the HCII and PLSS.

My family has been responsible for a large part of my intellectual development, and I thank my parents and sister for a lifetime of support. I also greatly appreciate the help my wife's parents have provided during the past few years.

Of course without the love and encouragement of my wife Jessica this work would have never been completed.

Table of Contents

List of Figures and Tables

Study 1: Handwriting and Text-Editing

Study 2: Copy-Pasting

Study 3: Intervening on Selection

Designing Optional Interventions

Study 4: Restricting Selection

Designing Highlighting

Study 5: Highlighting vs. Copy-Paste

Chapter 1: Background

The central question motivating the research described in this thesis regards how note-taking changes in a digital environment. While studies have shown that technology affects how students take notes, there is a dearth of literature regarding the impact these changes in behavior have on the learning gains realized by traditional note-taking. My research attempts to begin to address this deficit by looking at how specific interface decisions affect both behavioral and learning outcomes.

These outcomes must be placed within the rich theoretical and practical history of note-taking research. As the work focuses on note-taking while reading, interpreting the results relative to theoretical models of reading is also useful. In the following section, I will review what is currently known about traditional note-taking, including note-taking behaviors, learning outcomes, and the links that have been found between the two. I will also describe current note-taking technology and argue that its limited evaluation reveals the need for deeper analyses of the type described in this thesis.

Note-Taking as an Educational Tool

Note-taking is an important student behavior. As many as 99% of students take notes during lecture (Palmatier and Bennett, 1974). Studies have indicated that between 71% (Palmatier and Bennett 1974) and 91% (Lonka et. al. 1994; Fowler & Barker 1974) of students take notes while reading. Up to 96% of students believe that note-taking is an important part of their educational experience (Palmatier & Bennett 1974). Students report using note-taking to accomplish a variety of goals, including learning, maintaining attention during lectures and directing how they study (Van Meter et. al. 1994). Though formal strategies of note-taking exist, as few as 17% of students report having received training regarding how they should take notes (Palmatier and Bennett, 1974).

Note-taking research has historically concentrated on note-taking practices in the context of both lectures and reading. The results are similar enough to be described together, though throughout this review I will explicitly discuss aspects specific to each domain. With regards to learning, note-taking benefits have been placed in two categories: **Encoding** and **External Storage** (Carter & Van Matre, 1975). Encoding benefits are accrued through the act of note-taking. The act of recording an idea in notes facilitates learning, regardless of whether the notes are later reviewed. External Storage benefits are derived from students reviewing their notes. In this case, notes are useful as documents that can be reviewed prior to tests.

Evidence exists for both encoding and external storage effects, though it is not conclusive. While some studies find both effects (Rickards & Friedman 1978), others find only external storage benefits (Carter & Van Matre 1975), while still others find that encoding benefits only occur when students take notes knowing they will be able to review (Kiewra et. al. 1991). Kobayashi conducted meta-analyses of studies addressing both encoding and external storage. He found moderate encoding benefits (2005), and strong external storage benefits (2006). These analyses included many moderator variables, some of which will be discussed in further detail below. It is interesting to note that students are conscious of both benefits of note-taking, and report both as reasons for taking notes (Van Meter et. al. 1994). It also appears that knowing they will not be able to review their notes reduces the amount of notes students take (Slotte 1999).

Before speaking in greater detail about learning results, it is important to point out the importance of time-on-task with regards to research on note-taking while reading. Many studies have found that taking notes while reading increases the amount of time students spend with the material (e.g. Bretzing & Kulhavy 1979; Dyer et. al. 1979; Rickards & Friedman 1978). However, different note-taking techniques can take different amounts of time, while producing the same learning results (Annis 1975; Annis 1978). This difference in **learning efficiency** has been an

important factor in my own research, which has found that interfaces affect time-on-task, even where learning remains unchanged.

Technology, Note-taking, and Learning

Technology is playing a greater role in education than in years past. In 2004, the National Center for Education Statistics found that 56% of two and four year degree granting institutions had distance education offerings. An additional 12% had a plan to offer distance education in the following three years. Ninety percent of these programs used asynchronous online technologies to implement distance education (Livingston & Wirt 2004). All indications are that this trend of integrating technology and education will increase.

There are several arguments that can be made regarding the need for note-taking applications in these digital environments. It is reasonable to assume that as note-taking has been found to promote learning in traditional settings, it will do so online or in digital textbooks. At the very least, this possibility should be explored. In addition, students appear to want the ability to take notes on digital materials, and this unmet need may get in the way of other beneficial aspects of online learning environments. One study investigating the effectiveness of an online learning environment found that students were printing out online modules in part so they could annotate the printouts (Scheines et. al. 2005). In doing so, students missed out on the opportunity to complete interactive online materials, which were associated with superior learning outcomes.

A final argument in support of the evaluation of note-taking interfaces is that a variety of note-taking interfaces have already been built for educational purposes without satisfactory research regarding their behavioral and educational outcomes. I will discuss several of these tools below. Their focus is often on developing new note-taking techniques, rather than achieving the benefits of traditional note-taking. Technology is changing how note-taking can be supported, and many developers are taking advantage of this opportunity. As mentioned by Wolfe and Neuwirth, notes and annotations written in text can be shared just as they were in the early days of

books (2001). It is no longer defacement because it is not permanently attached. Many note-taking devices have been built to allow for discussions grounded in notes or annotations. Other devices have been built to allow multiple students to simultaneously access a single notepad during lecture.

There remains a great deal of basic research to be done. If a note-taking application is built for educational purposes, it should at the very least be designed to maintain the learning benefits achieved through traditional note-taking. However, few studies even evaluate note-taking behavior systematically, concentrating instead on satisfaction and motivation. In some cases they simply find digital note-taking to be cumbersome. As will be described below, those that do evaluate behavior often find that the way students take notes changes.

Though behaviors change, research has not investigated the impact these changes have on learning outcomes. If we want to build note-taking interfaces that serve educational goals, we need to do more than explore the boundaries of what technology can support. It is essential that we map functionality to behaviors, and behaviors to learning outcomes as well as satisfaction outcomes. In order to do so, we first must understand more about what we already know about how note-taking impacts learning, and what behaviors are involved.

Cognitive Processes

The research described in this thesis involves mostly manipulation of the techniques students use to record notes. As such, it deals mostly with the encoding benefits achieved through note-taking. Most of the theoretical literature also addresses the underlying causes of the encoding effects of note-taking rather than review effects. It is important to keep in mind, however, that changes in processes may affect external storage benefits, as the notes from which students review change. There is very little literature regarding what constitutes effective notes for review. I will first discuss the broad hypotheses regarding what underlies the encoding benefits of note-taking, and then describe what is known about the connections between note-taking and learning.

Two hypotheses have been developed addressing the mechanisms by which note-taking promotes learning: focusing and elaboration. In the note-taking literature these have often been referred to as "attention" and "generation" respectively (e.g. Peper & Mayer 1986). I choose the reading comprehension language of Reder, as I will also use its theories to interpret my data and the note-taking literature at large (Reder 1985).

The **focusing hypothesis** states that note-taking increases encoding when it increases the attention students give to the learning material, requiring them to focus on critical information. In other words "Note-taking forces learner to pay more attention to the presented material" or to "process the material more deeply." The **elaboration hypothesis** is that note-taking promotes learning when it is a generative activity, connecting multiple knowledge components. This requires that "additional cognitive processes are involved, for example the degree to which the learner is able to actively relate the material to existing knowledge." (both quotes Peper & Mayer 1986) This may involve generating links to prior knowledge, or even connecting distinct concepts within the learning materials.

Ideas of focusing and elaboration fit well with Kintsch's prominent levels of comprehension model of reading comprehension (1994 for summary). The most basic level of comprehension in this model involves the development of a **surface** understanding of the words, phrases, and linguistic relations. The second level involves the creation of a **textbase**, which is the reader's representation of the semantic and rhetorical structure of the text itself. The textbase does not go beyond the structure or information within the text itself. As we will discuss below, focusing is most likely to affect reading comprehension at this level. The last level is the **situation model**, in which the reading is elaborated and integrated, most often with prior knowledge. Elaborative behaviors would be most likely to strengthen this level. While a good textbase will allow a reader to summarize texts and answer questions about content, according to Kintsch a situation model is required for inference and problem solving test items.

Focusing and Attention

While students report using note-taking to avoid distraction (Van Meter et. al., 1994), the question remains as to how note-taking focuses students' attention. When students record notes, they are spending more time rehearsing the ideas being recorded. Such rehearsing of reading material results in improved memory (Reder 1985). Note-taking may also encourage students to identify the critical components of the idea being recorded, which can be seen in the wordiness with which an idea is recorded. Reduced wordiness would also reduce short-term memory load. So note-taking may encourage students to spend more time rehearsing the most critical components of the ideas they are recording, which would increase the chances the idea will be strongly linked in the textbase (Kintsch 1998).

Note-taking may also help students identify the key ideas within the learning material. As Johnson states (1998), note-taking "improves retention of passage material when it focuses students' attention on identifying ideas of high structural importance." Some studies have even found that remembering recorded ideas helps students reconstruct higher level ideas not recorded (Rickards & Friedman 1978, Johnson 1998). By increasing focus on structure, note-taking is helping students strengthen the rhetorical component of the textbase.

Elaboration and Coordination

The elaboration hypothesis can take two forms, though evidence is limited for both. The first and most popular form of the hypothesis is that note-taking is beneficial when it causes students to elaborate on the ideas they record by connecting it with their prior knowledge. It is often thought that notes students record in their own words are superior to notes recorded verbatim from the learning material, as they are a sign of elaboration. However, as we will see below, learning research does not conclusively support this hypothesis. Rewording notes does not require the integration of outside knowledge or an augmented understanding of the structure of the learning material. Simply putting notes in ones' own words does not require the elaboration of a situation model, as it does not necessarily involve a connection with outside knowledge. Still, it may be that note-taking is a bridge that allows students to

connect their learning with prior knowledge. In their studies, Peverly et. al. found that background knowledge only affects learning outcomes when students take notes (2003). Note-taking may help students elaborate what they are reading by connecting it with what they already know.

Note-taking may also improve the situation model by encouraging **bridging inferences**. These are inferences that do not require outside knowledge, involving instead the coordination of ideas located in separate sections of text. A student's notepad may facilitate these types of inferences by serving as a record of ideas from previous pages that would not otherwise be available. Some note-taking techniques, such as typing or handwriting, create a separate representation that could be referred to on subsequent pages, whereas others, such as highlighting, do not create notes separate from the material itself. While the former would promote bridging inferences, the latter would not. Such simultaneous availability of multiple learning sources has previously been shown to improve learning outcomes such as essay writing (Wiley 2001).

Some evaluations have implicitly tested this elaboration hypothesis by comparing handwritten notes with highlighting. One study found no learning difference, but did find that highlighting decreased time on task (Annis & Davis, 1978). Another found that handwritten notes produced improved learning relative to highlighting (Kulhavy et. al., 1975). However, in this case handwritten notes were recorded on the learning material, which eliminates coordination. Neither study made it clear how many pages students read. It would be expected that coordination effects would be more salient with longer material, and the passage length of these studies was 1525 and 845 words, respectively, meaning little coordination was possible. It would be more useful to test this hypothesis with in larger passages whose structure was designed to require coordination. The studies in this dissertation involve materials containing over 9000 words across 15 pages.

While bridging inferences involve structure not explicit in the text, it is not clear this fully qualifies as "something beyond text" (Kintsch 1994) required by Kintsch's

situation model. In much of his writing, Kintsch expresses the dependence of the situation model on prior knowledge (e.g. 1998). However, in others it is sufficient "that information from different places in the textbase be combined" (Mannes & Kintsch 1987). In another study items are included that only require bridging inferences, but are interpreted relative to the situation model (McNamara 1996).

Summary of Theory

Focusing and Elaboration were introduced as two hypothesized means by which note-taking encourages learning. With regards to focusing, note-taking could encourage students to both identify key ideas and rehearse the critical components of the key ideas being recorded. This would strengthen their textbase, which in Kintsch's model is the student's model of the semantic and rhetorical structure of the text. Focusing could be seen in both the recording of a note, and the wordiness with which it is recorded. However, this strengthened textbase would only facilitate recall, and note-taking has been shown to increase performance on problem-solving and inference test items. These are under the purview of the situation model, which goes beyond the information contained in the learning materials.

The situation model is generally thought to require elaboration with background knowledge, which is not often seen in the notes students take. It may be that by strengthening the textbase, note-taking allows for this type of elaborative behavior. Another possibility is that note-taking increases the likelihood of bridging inferences by allowing the note-taker to view information from multiple pages simultaneously. If this is the case, note-taking interfaces that provide visible notepads would promote learning to a greater degree than others, such as highlighting interfaces, that do not include a notepad.

These questions are especially important considerations with regards to how note-taking should be supported in a digital environment. Input techniques are far different, which could have an effect on note-taking outcomes. Typing is faster than handwriting (Card et. al. 1983), and copy-pasting is faster still. Does the time cost of note-taking with pencil and paper contribute to focus benefits? On the other hand,

as we will discuss later digital annotations are easily extracted from the learning content, so they can be treated as either anchored or unanchored depending on the design and learning considerations. Would extracting annotations allow students to use their notes for more elaborative tasks? This thesis spends a good deal of time evaluating the relationship between interfaces and focus, and includes a study that begins to address the benefit of extracting annotations.

Note-Taking Behaviors

While the focusing and elaboration hypotheses have not seen detailed evaluation in the note-taking literature, it is possible to gain insight by evaluating the contents of notes students produce. In order to do so, students' notes must be coded with regards to the ideas contained within. These content analyses are somewhat underrepresented in the literature (Kiewra et. al. 1984), but is made easier when notes are collected in digital format. Here I will discuss the behaviors that have been addressed most thoroughly in the literature, though the evidence is not conclusive for any particular behavior.

Verbatim Notes

Notes are often transcribed word for word from either lectures or readings. As Carter points out, "note taking is more likely to resemble verbatim transcription of the sort that occurs with copying frames in programmed instruction, than the more beneficial elaborative activities associated with meaningful learning" (Carter & Van Matre, 1975). Implicit is the commonly held assumption (see also Kiewra et. al. 1985) that verbatim note-taking does not promote learning because it is not elaborative.

The literature and this thesis contrast notes taken verbatim with notes taken in students' "own words." It is important to mention that no coding of "own words" in the literature requires the inclusion of outside information. While the most stringent definitions require the use of different words or paraphrasing, none require the presence of ideas that are not in the learning material. As I will argue later, paraphrasing is not necessarily elaborative, and there may be other benefits of verbatim note-taking.

Verbatim note-taking is quite common. One study found that sixty percent of notes taken while reading were recorded in a verbatim format (Bretzing & Kulhavy, 1981). Interestingly, another study found that while students took more notes in their own words than they did in verbatim format, when they knew they were not going to review students increased the amount of verbatim notes they recorded (Slotte & Lonka, 1999). Others have found verbatim note-taking to be a persistent behavior. Even when instructed to take notes in a non-verbatim format, students continue to record their notes in a verbatim format (Kiewra et. al 1984), and thus receive no learning benefit. Students report having clear strategies for taking verbatim notes, using them to record definitions and to help them review when they expect tests of recognition (van Meter et. al., 1994).

The implications verbatim note-taking has for learning are unclear. One study of note-taking during lecture found that students performed better if they were forced to take verbatim notes, especially if the notes were not reviewed (Carter & Van Matre, 1975). However, when Peper and Mayer compared students taking notes in their own words with students taking verbatim notes (1986), they found that students who took notes in their own words performed worse on immediate recognition tests, but better on problem solving and far transfer, though of course this result is subject to selection bias. They interpret this as evidence for elaboration, which they say interferes with performance on verbatim recognition questions. Though this also implies the construction of a situation model, it is not clear that notes written in students own words is elaborative in the Kintsch sense, as the textbase allows students to summarize or reword ideas from the learning material (Kintsch, 1994).

More studies examine the value of verbatim note-taking while reading. One study found that requiring verbatim note-taking produced better learning outcomes than paraphrased note-taking (Quade, 1996). In contrast, Bretzing and Kulhavy found that verbatim note-taking was worse than summarized (creating summaries of passages) and paraphrased (paraphrasing the ideas recorded) notes (1979). They

believed that verbatim note-taking promoted skimming passages, while summarizing and paraphrasing required students to process the material more deeply. They later found that verbatim note-taking actually performed better for formally worded texts, whereas paraphrased note-taking performed better for informal texts (1981). It may be that text difficulty plays part in the effectiveness of verbatim note-taking.

We know verbatim note-taking is a common and persistent activity. While many researchers believe it to be a negative behavior, the evidence is equivocal. Rewording may not always be an elaborative behavior, as it may simply involve using new words to express the same idea, rather than adding additional information to the idea being recorded. My own research has shown no negative effects of verbatim note-taking, though it does provide some evidence linking it with a potentially negative skimming activity. As we will see, technology can be used to manipulate verbatim note-taking strategies in ways that may allow us to understand this issue in greater depth.

Wordiness

Another debate exists regarding the wordiness with which students should record their notes. Wordiness is calculated as the total number of words in notes divided by the total number of ideas. It can also be classified on a per idea basis as the number of words used to express a given idea. Some researchers believe wordier notes are indicative of increased elaboration of the learning material, and should be encouraged. It also may be that wordiness represents more time spent rehearsing the ideas being recorded. On the other hand, there is some evidence that it is important to be more efficient in recording ideas. Efficiency may reduce cognitive load, and increase the time available for rehearsal. This may be especially important in lectures, where time spent note-taking may distract from learning opportunities. The evidence is mixed, and probably relies on both presentation format (lecture or reading) and how the notes were recorded.

An early study of note-taking efficiency in lecture found a significant positive correlation between less wordy notes and recall (Howe 1970). Another study found no correlation between wordiness and learning, only between the quality of notes

and learning (Fisher 1973). After having found both the number of ideas recorded and the wordiness with which they are recorded to be positively associated with learning, Kiewra evaluated matrix and outline note-taking strategies, which he expected to increase both measures by identifying the material students should record (Kiewra 1987). These will be discussed further below. Most of his studies found matrix note-taking produced wordier notes in lecture (Kiewra et. al., 1989, 1995), but did not appear to produce wordier notes while reading (1989). As we will see below, matrix notes often do perform better on learning outcomes in lecture than non-matrix notes (e.g. Kiewra 1995), but it is not clear if this is due to increased wordiness or the students being directed to focus on key ideas in their note-taking activities.

The value of wordy notes is unclear, especially with regards to reading, where there little work has addressed the issue. With regards to technology, Van Oostendorp has found that taking notes on paper while reading online material is more efficient than taking notes on an integrated text editor, but did not observe a difference in learning outcomes (1996). My own studies have found that the impact of wordiness depends on the interface used to record notes. When there was no cost associated with the production of wordier notes, wordiness was negatively associated with learning. Wordiness is another feature that can be manipulated in note-taking applications.

Presence of Ideas in Notes

The presence of ideas in notes has been associated with learning since Crawford's early studies (Crawford Nov 1925), where it was found that an idea recorded in notes during lecture was more likely to be recalled than an idea not present in notes. Other lecture-based studies (e.g. Howe) have replicated this result, while others have found positive correlations between the total number of ideas recorded in notes and learning outcomes (Fisher 1973). Reading studies have found similar results (e.g. Bretzing 1981). One study found that if students do not reread the material, ideas that are not present in notes are more likely to be missed when tested (Dyer 1979). Another found that students were likely to free-recall ideas in the same order as they were recorded in the students' notes (Schultz, 1972).

The presence of an idea in notes appears to be important, especially when students cannot review the material. This raises an interesting question regarding what information is important to record. It is clear that students should record ideas that they will be asked to recall. Recording additional ideas that supplement the important points may be useful. However, there may be a crossover point, where increasing note quantity reduces the effectiveness of note-taking. Though this question has not been directly addressed in the literature, many studies restrict the amount of notes students can take. The actual restrictions have no empirical basis. My studies indicate technology can increase the amount of ideas students record, especially with regard to peripheral ideas. Note quantity can be manipulated fairly easily in note-taking applications, though that manipulation will not be explored in this thesis.

Individual and Content Factors

There are a variety of individual traits and attributes of the learning material that have been implicated in note-taking outcomes. *Signaling* involves emphasizing important concepts being instructed. In lecture this is most often done through words (e.g. "this is a key point") and pauses while in reading this is done through words, headings, markups, callouts and the like. For lecture, it has been found that note-taking is useful when signaling is used, especially for field dependent learners (Rickards 1997). While many would suppose lecture speed is an important factor in the viability of note-taking, the limited amount of current studies do not support this hypothesis (Peters, 1972). Signaling has also been shown to be important for note-taking while reading (Fox, 1985). Many of the important concepts in the materials used in my own studies are signalled by headers.

Age appears to be an important individual attribute of note-takers. Younger note-takers are less confident in their note-taking abilities, and those who are less confident perform worse on learning outcomes (Carrier et. al., 1988). Kobayashi's meta-analyses showed that younger students are more likely to be affected by note-taking interventions (2005, 2006). Schellings points out that younger students have immature note-taking strategies (1995). Older students may thus be more resistant to

note-taking interventions. Students appear to have developed strategies after their first year in college. My studies involve students of college age, ranging from freshman to graduate students.

Younger students also show a high degree of variability with regards to their note-taking strategies and what they consider important to record. Interestingly, not only are students not consistent with each other, instructors of the same content do not identify the same ideas as important to note (Schellings & Van Hout-Wolters, 1995). It appears note-taking is a highly individualized process.

To a certain degree, the personal nature of note-taking practices may account for findings regarding whether students should take their own notes or professors should provide lecture notes to students. In general, the literature has found that note-taking requires "active participation for optimum results" (Fowler & Barker 1974). Students perform better if they can review their own notes (Fisher & Harris 1973), and perform better when they are left to their own devices (Rickards & August 1975). Finally, they perform best when not expecting any specific type of test, for instance multiple choice or essay (Kulhavy et. al. 1975).

Interventions

While the topics described above suggest several interventions that will be explored in this dissertation, a limited range of interventions have been evaluated in the literature. Only matrix note-taking, which I will describe in more detail below, has been explored systematically. Most interventions have been evaluated in one-off studies.

Widespread prescriptions for effective note-taking exist, but little empirical data exists supporting the educational effectiveness of any of them. In the seminal "How to Read a Book", Adler and van Doren (1972) lay out note-taking guidelines and strategies for reading. In their view, note-taking promotes active reading, with "the pencil [becoming] the sign of your alertness while you read." It recommends a variety of note-taking techniques, including underlining, lines at margins, and circling

words. These practices have been observed in more recent studies of note-taking behavior (e.g. Marshall 1998), and digital devices have been explicitly built to support 'active reading' as defined in this work (e.g. Schilit et. al. 1998).

While Adler's system provides recommendations for reading, the Cornell Note-taking System is the most popular prescription for lecture-based note-taking, and is also recommended for use while reading (Pauk, 2000). It includes structural recommendations for organizing the paper on which students take notes, strategies for summarizing, rewriting, and reducing the material students are processing. Again, these recommendations have not been empirically validated. Many other instructional materials regarding how to take notes also include the recommendation to put notes in students' own words, which as we have seen and will see in my own work, is not an entirely substantiated recommendation.

There are several examples of one-off studies of other note-taking interventions. Rickards found that students who were instructed to underline ideas of high structural importance recalled fewer details than students who were instructed to underline what they wished, and recalled an equivalent number of ideas of high structural importance (Rickards & August, 1994). Another study instructed students to take one of three types of notes during lecture: conceptual (only record main ideas and attempt to summarize), relational (link main ideas to own knowledge) and factual (just record details). There were no learning differences, and the interventions were somewhat ineffective, as students reverted to a verbatim note-taking style in all conditions (Kiewra & Fletcher, 1984).

Note-taking appears to be somewhat resistant to intervention. The above studies were conducted with college students. As mentioned above, many already have fully developed their note-taking strategies. Kobayashi's meta-analyses indicate that pretraining and instruction are not effective note-taking interventions. Kiewra's study shows that students revert to their standard practices. Behavioral interventions, on the other hand, are effective with regards to learning (Kobayashi 2005, 2006). With behavioral interventions, students are forced to take a specific type of note. One

benefit of technology may be the ability to intervene in ways that could previously only be instructed or trained, or even give in-process feedback rather than pre-training.

Graphical Organizers

The most common behavioral intervention involves providing students with graphical organizers to use for note-taking. Most commonly studied is matrix note-taking, in which students are given tables with rows and columns indicating attributes or items. Students then fill the cells with specific information that matches the attribute indicated for the given item. In a graphical organizer for a geography topic, for example, the rows could represent countries, while the columns represent features, such as climate or total rainfall. Other graphical organizers include partial outlines displaying the attributes and items in hierarchical form.

Kiewra designed the early graphical organizers in order to increase the amount of notes taken, and the overall coverage present in students' notes. An early study (Kiewra et. al., 1989) found that matrices and outlines produced more notes than traditional note-taking, but did not have any effects on learning for either lecture or reading. A later study did find matrix notes to be better than traditional note-taking (Kiewra & Dubois 1991). In his later work, Kiewra concluded that in lecture outlines should follow the lecture presentation order and matrix notes should be rigidly organized according to hierarchy, not lecture (Kiewra 1995). While both increase the number of ideas, and efficiency with which they are expressed, only the flexible outlines produce better learning results than traditional note-taking. A study comparing outlines with graphical organizers for reading found that partially completed graphical organizers are better than outlines for application questions (Katayama & Crooks, 2001) for online learning materials. This study did not include a traditional comparison condition.

Graphical organizers and outlines are structured note-taking interventions, requiring students to record specific ideas in a very specific manner. This is opposed to unstructured note-taking, in which students are free to record what they deem most

important. Interventions in unstructured note-taking manipulate how students take notes, not what they record. The work in this thesis evaluates unstructured note-taking for several reasons. First, as mentioned above students do not all record the same information while taking notes, and freedom to take notes appears to be a useful feature. Graphical organizers tell the student what they should be recording in their notes. Students may perceive these as another assignment or task, and approach the task in a different fashion then they would when note-taking is internally motivated. In addition, graphical organizers and outlines make demands of the learning material and the instructors. They require very well structured learning content. Even highly structured material may require multiple graphical organizers when the text is complicated (e.g., Robinson & Kiewra, 1995). This requires additional work on the part of the instructor not required in less structured methods of note-taking.

Restrictions

Many studies of note-taking in the context of reading intervene on the note-taking process by restricting the number of notes students can record. The reasoning behind restrictions is fairly straightforward. Johnson states "certainly, the amount of underlining must be controlled… otherwise a few subjects underline everything or underline nothing" (1988). However, there never has been any empirical evidence that eliminating restrictions leads to too much note-taking, though part of the problem is that it is unclear how "too much" would be defined. For example, though one study found notes were taken on 31% of passages (Rickards & Friedman, 1978), it was not clear whether this was too little, appropriate, or too much, and no link was observed between quantity and learning outcomes.

Limits have been placed in much of the literature. They range from allowing users to take a note on or highlight 1 sentence per paragraph (Rickards & August, 1975) to three lines a page (Bretzing & Kulhavy, 1979; Kulhavy et. al., 1975; Dyer et. al., 1979) to seven lines a page (Bretzing & Kulhavy 1981). However, these restrictions have rarely been compared with each other, or to unrestricted notes. One study found that restricted verbatim note-taking was worse than unrestricted paraphrased notes

(Bretzing & Kulhavy 1979). I am only aware of one instance in which restriction itself was manipulated. This study comparing restricted to unrestricted note-taking found that restrictions were better for remembering details, but unrestricted notes were better if students were allowed to study their notes (Santa et. al., 1979). Students actually performed worse than no notes for main ideas, but better for detailed ideas. This contrasts with a study where restricted underliners recalled main ideas better than students who did not take notes (Johnson, 1988).

The impact of restrictions on note-taking is unclear. Though it may have some relationship with identifying structural elements in the learning material, this appears to depend on how students record their notes. It is also unclear how these restrictions could be enforced for paper documents outside of the laboratory. However, if students are studying and taking notes using digital appliances, implementing these restrictions can be quite facile. The studies reported here indicate restrictions may be an important area of future investigation, as it appears some of the affordances of technology lead to disinhibited note-taking behavior.

Note-Taking Technology

In the following sections I discuss the current state of note-taking technology. I first describe learning technologies, and then discuss more general-purpose applications.

Standard Learning Interfaces

Note-taking applications built for educational purposes often resemble text-editors embedded into the learning material. These are usually implemented in frames, with one frame containing the learning material and the other containing the note-taking application. These text-editors function as unstructured note-taking applications, much as traditional pencil-and-paper note-taking. Though important differences exist, typing and handwriting provide the freedom to use any words and organize and edit notes as the user sees fit. In addition, both involve the creation of separate documents. Differences include an inability to draw or create tables, though these features have not been evaluated in the note-taking literature. Though such devices

could include features such as copy-paste, in general they only support text-entry through typing.

Studies have been conducted regarding both the learning outcomes associated with such tools and the note-taking behaviors produced. One study found that students using a text editor performed equivalently to students using paper on review tests, and both conditions performed better than no notes (Quade 1995). Another study analyzed note-taking behavior, finding that more notes were produced using a text editor than were recorded on paper, and that the notes were wordier. No learning differences were found. (Van Oostendorp 1996). Interestingly, verbatim notes were related to better learning results than were notes recorded in students' own words, perhaps because they were able to review accurate transcripts of the learning material. In a study in which students could choose to take notes on two paragraphs using either paper or a text-editor that did not allow copy-paste, no learning differences were found (Rice, 1994). Another study found that if students were forced to take notes using a text-editor, they performed better than if the tool was optional. Both the required and the optional tool performed better on learning outcomes than the no-notes condition (Armel, 1995).

Several studies have compared applications that allow students to highlight digital text with the highlighting of text documents. One study found no learning differences, but did find that on paper more rhetorical units were highlighted, whereas on computer more idea units were highlighted (Rice, 1994). However, participants in the experiment chose their own condition (paper or computer), so there may have been other factors involved. In a controlled study of the use of highlighting for editing documents, it was found that error catching was equivalent on paper and computer (van Oostendorp 1996). This was not, however, a measure of learning.

These studies provide a first level analysis of the relationship between unstructured note-taking in the digital and physical worlds. My own research involves

manipulating such interfaces in order to explore note-taking in greater depth and guide the design of future unstructured applications.

Graphical Organizers

The most mature example of evaluations of note-taking applications follows the graphical organizer tradition of the paper-based note-taking literature. Only here do we see evaluations of the effect of manipulating specific interaction techniques. Results may not generalize beyond structured note-taking, as these studies do not compare online behavior with physical notes or unstructured note-taking.

An early study found that partially filled in online graphical organizers produced better outcomes on immediate learning tests than did complete ones (their completeness meant students would not use them to record notes), and that students using them forgot less (Katayama & Crooks, 2001). A subsequent study found that although students preferred using copy-paste to typing in the partial graphical organizers, they performed better on tests when they typed (Katayama et. al., 2005). Another study found that restricting the amount of text a student could copy-paste improved test performance (Igo et. al., 2005).

I have described the limitations of structured note-taking above and will not be using them in my thesis. However, studies of graphical organizers do provide interesting examples of how specific functionality such as copy-paste can be manipulated and how such manipulations can affect not only how students take notes, but how much they learn. Again, results may not generalize to unstructured note-taking applications. Structured note-taking provides students with additional information in the form of headings. Restrictions may have different effects when students are not guided as to what ideas to record and are not given the additional information to supplement the restricted notes. These problems are in addition to the ones described above; namely, graphical organizers require highly structured learning material, additional work from the author, and may be substantially different from what students perceive to be note-taking.

Other Educational Devices

A variety of more feature-laden note-taking applications have been built but not evaluated with regards to learning. While they generally mimic traditional procedural features of note-taking they often add new interactive components. While the above tools were created for reading materials, the tools described here are often built for lecture.

These lecture-based tools often allow for a variety of free-form note-taking, using a "digital pen" metaphor. Stupad, created for Georgia Tech's Classroom 2000 project is one example of a freeform note-taking device on a tablet platform. Students were given digital copies of the slides presented in lecture, which they could write on using a stylus. Students were found to create verbatim transcriptions of what the professor wrote during class, which was taken by designers to be a negative behavior though as mentioned above this is not clearly supported by the literature. In addition, the tablets were perceived as unwieldy and distracting from the lecture (Abowd 2000). A later redesign effectively reduced the verbatim transcriptions by integrating live captures of what the professor was writing (Truong et. al., 1999).

Livenotes is another example of a tablet-based note-taking application developed for lecture (Kam et. al., 2005). This pen-based device promoted cooperative note-taking, where a group of students created one set of notes together. This procedural design was based on concepts of collaborative learning. Though the number of subjects in the learning study was too small to produce statistically significant results, behavioral effects were found. The collaborative note-taking process was seen to increase commentary and discussion while reducing the recording of individual ideas.

Educators have also built annotation based note-taking devices for reading (Mason et. al., 1999). These devices allow students to highlight text by selecting it with the cursor, much as a physical highlighter marks up a paper document. Many of these devices go one step further, allowing students to attach typed comments to any individual highlight (Lebow et. al., 2006). These highlights and comments are then

often used as the basis for online discussion between students. Though the learning outcomes of these features have not been evaluated, there is evidence that the features included in such annotation devices alter how students take notes. For example, sharing results in more formally worded comments than does writing for oneself, and the requirement to attach comments to specific highlights reduces the tendency to make general commentary (Marshall & Brush, 2004).

Non-Educational Devices

There are a variety of note-taking applications built for non-educational purposes. These devices have most often been "designed to enhance the traditional paper note-taking activity rather than define a new process as personal organizers and portable computers do" (Wilcox et. al., 1997). The real traction is not seen to be in changing how notes are recorded, but in supporting the act of reviewing notes. Here technology can take advantage of the best aspects of multiple types of notes such as individual pages, note-cards, and notebooks (Schilit et. al., 1998). To a large degree these tools are designed to address the question of why, in an age of technology and digital data, paper remains a focal point of writing work (e.g. Adler et. al., 1998).

Examples of changes to the note-taking process do exists with respect to small devices such as PDAs, where traditional handwritten stylus-based input can be extremely limited for large amounts of text by screen size and the lack of a keyboard. Goals of these devices here include increasing the speed at which text can be input into the device (Ward et. al., 2000) or sharing meeting notes (Davis et. al., 1998)

In general, support for the process of recording notes is inspired by observations of how people record notes on paper. The seminal work on this was conducted by Marshall, who participated in several of the projects described above. She categorized the purpose of note-taking on continuums from formal (metadata) to informal (marginalia) and explicit (written text) to implicit (highlights or other emphasis). She identified a variety of methods of taking notes, including lines and arrows associating elements, emphasis marks such as highlights, resegmentation, written notes, and categorizations (e.g. using color) (Marshall 1998). Much of this

work was mirrored in the prescriptive text by Adler described above (Adler & Van Dorren 1972).

Summary and Goals

A large body of literature on traditional note-taking has shown it to be an effective educational activity. Not only are positive learning outcomes achieved from having notes to review, but the very act of note-taking has often been shown to improve learning. Assessing the quality of students' notes may also be useful, as learning has been associated with the types of notes students take. It may be easier to do this quality assessment with digital notes, as they can be automatically collected and analyzed. However, technology has also been shown to change the way student take notes, and it is important to understand how these changes affect learning gains achieved through traditional note-taking if at all.

Technology is also promising with regards to intervening in the note-taking process. Training students or giving them instruction regarding how to take notes does not appear to be effective at changing behavior or improving learning. Only behavioral interventions, that is forcing students to take notes in specific ways, have been shown to effectively increase performance on learning outcomes. Even here it is unclear how the interventions would be enforced outside of the laboratory. However, technology allows us to build the very system with which students take notes, allowing for a control over the note-taking process that can persist in the real world.

New methods of note-taking can be supported in digital environments. I have described above systems in which students use annotations as the basis for discussion and one system in which students collaboratively take one set of notes. My research deals with note-taking environments that map more closely to traditional note-taking in order to provide results which can then inform the more novel technologies. It starts with text-editor based note-taking, due to its similarity to notes taken using pencil and paper, which form the basis of the majority of the previous literature.

This work evaluates the way features of note-taking applications affect behavior, and how differences in behavior affect performance on learning outcomes. It also investigates the design of interventions intended to encourage students to take notes in ways that are associated with learning gains. It focuses on note-taking techniques where students select text using the mouse in order to record it in their notes. Initially this focuses on copy-paste based note-taking, but later in this thesis I will explore similar highlighting interactions. The following pilot study describes how my attention was first drawn to such **selection-based note-taking**.

Chapter 2: Handwriting and Text-Editing

A pilot study was conducted to identify differences between traditional pencil-and-paper note-taking and note-taking online using a text-editor. As described above this comparison has been done previously in (Rice 1994, Quade 1995, Armel 1995, Van Oostendorp 1996). These studies were somewhat limited with regards to the quantity of text (as short as two paragraphs). While they found that a text-editor produced a greater quantity of notes, they did not make it clear what was responsible for this increase in quantity. The main purpose of this study was to elucidate the functionality that produces the differences seen in note-taking behavior, in order to identify features and behaviors that should be investigated in greater depth.

Overview

As mentioned above, this research deals with unstructured note-taking applications, which allow students to record the ideas they find most important, rather than a structured note-taking application such as the graphical organizers described above, which direct students to record certain ideas in a specific format. These structured devices require additional work from instructors, give the students additional information, and guide the note-taking process. Though these are the most thoroughly studied note-taking applications, note-taking practices in general, and note-taking applications, are dominated by unstructured note-taking. There exists a broader range of unstructured note-taking applications, from handwriting to typing to annotation and highlighting. Studies of unstructured note-taking should generalize to a broader range of applications.

The initial studies were aimed at understanding the key interface elements involved in a common unstructured note-taking application, and the behaviors affected by these elements. An embedded text-editor (see Figure 1) was chosen for analysis, as it would allow students the option of either recording notes in their own words or verbatim, much as they would be able to with pencil and paper. Other unstructured

interfaces, such as highlighting, would only allow verbatim note-taking. In multi-featured applications, such as ones that allowed for comments to be attached to highlights, it would be difficult to attribute any observed effects to any specific feature. Practically, text editors were likely to produce interesting results, as previous studies had found behavioral differences, though they had not identified the causes.

Hypotheses

I developed several hypotheses regarding the differences between note-taking using the text-editor and note-taking using pencil and paper.

H1: Students using the note-taking tool will take an increased quantity of notes. As mentioned above, this result has been observed in several other experiments. Several procedural benefits allow text-editing to be more efficient than handwriting. Handwriting is relatively slow when compared to typing speed. While handwriting speeds average about 15 words per minute, even typists who hunt-and-peck are capable of rates of 20-40 words per minute (Card et. al., 1983). A second speed related benefit comes from the ability to copy-and-paste material, which is even faster than typing.

The tool could also make fewer demands on attention than note-taking using paper. Students using this tool should be able to maintain focus on the materials at hand. Handwritten notes require the student to focus on the paper, while hands must be shifted from the mouse or keyboard to the writing implement and paper. An online tool allows students to stay focused on the monitor and keep their hands on the keyboard. Experienced touch-typists are able to refer to material and type simultaneously.

H2: Students using the note-taking tool will take more verbatim notes. Copy-paste by definition produces verbatim notes, unless those notes are then edited. In addition, it is possible that students occasionally take non-verbatim notes in order to save time by reducing the total number of words necessary.

<u>H3: Facilitated note-taking will lead to decreased encoding, and therefore poorer performance on post-tests.</u> This study was focused more on behavior than learning. However, we did include an immediate post-test taken directly from the course materials. If the text editor does in fact increase verbatim note-taking, we expect poorer performance on learning outcomes. As the research outlined above describes, verbatim note-taking is generally thought to produce poor learning, though the evidence is not conclusive. Note, however, that the note-taking technology studies cited above did not find this effect.

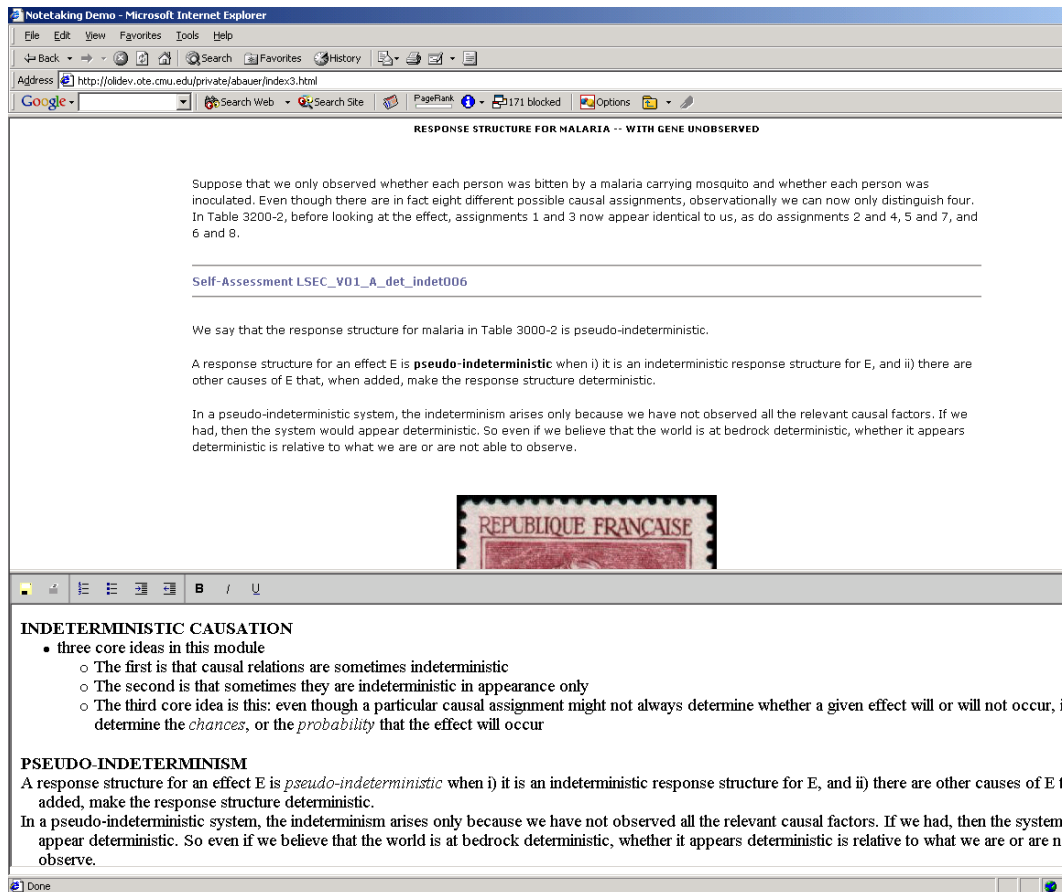**Figure 1: Note-taking interface used throughout this work. Implemented in frames, the content takes up the top two-thirds of the browser, and the note-taking application takes up the bottom third. The javascript text editor allows for basic outlining (bullet lists, indent, outdent) and text markup (bold, italic, underline), available through the toolbar at the top of the editor, or through standard Word-based keyboard shortcuts.**

Method

The experiment followed a within-subjects paradigm. Students took notes in two course modules using an embedded note-taking tool in one and paper in the other. Participants were tested individually.

Subjects

Fourteen students at Carnegie Mellon University were recruited for this study. Twelve were undergraduates, while two were graduate students. All had average to above average experience with word-processors. None used a "hunt-and-peck" typing strategy. No student reported being familiar with the course materials.

Experiment Design and Procedure

Students were told that they would be studying the second and third modules in an online course in Causal and Statistical Reasoning (Scheines et. al., 2005). They were asked to take notes while studying, and told that in an actual course they would use the notes for their own studying as well as for a weekly discussion section during which they would not have access to the online material. The note-taking assignment was counterbalanced so that the *paper-first* group used paper to take notes in the first module and the tool in the second, while the *tool-first* group used the tool in the first module and paper in the second. We were therefore looking for module by assignment interactions.

Participants were seated at a desk in front of a 17-inch monitor, keyboard, and a mouse. They were first asked to read a one-page summary of the course and the first module. For the paper condition, students were given the choice of using lined or blank paper, pens, pencils and highlighters. These were taken off the table during the module in which students were asked to take notes using the tool (see Figure 1 for an image of the interface). Prior to the tool condition, students were given a quick introduction to the tool. They were told it behaved like a simple word processor, and though they were not explicitly told they could copy and paste text into their notes, they were told that if they could paste images into their notes if they so desired.

Each module was expected to take approximately an hour to complete, as that was average amount of time students spent on a module in the actual course. A short multiple-choice quiz taken from the learning materials was given after each module. After completing the first module, subjects were given the option of taking a short break before starting the second module. Upon completion of the second quiz, students were given a survey. This survey included items regarding their favorite and least favorite features of the tools, whether they would use the tools in an online class, and features they would like in an online note-taking tool.

Dependent Measures

Several dependent measures were taken. Time to complete each module was recorded. Total number of words taken was recorded on a per module basis as a measure of the overall quantity of notes. The number of individual notes per module was also recorded. Finally, quiz scores were obtained after each module.

Number of notes was coded as individual participant entries into their notes. For paper they were defined by new lines whose disconnection from previous lines was made obvious by white space, indentations, or list entries. Horizontal white space greater than 3 times the average space between words also identified separate notes. Connecting arrows were also useful in identifying this type of note. As digital notes were recorded in HTML, they were much easier to code. New paragraphs and list entries indicated new notes. In practice this meant the note could be highlighted by double-clicking on the line.

Total amount of verbatim notes was also recorded. Verbatim notes are defined as all notes whose words are identical to ones used in the module. They include abbreviations and sets of words that are in the same order as those in the module but leave out only conjunctions, prepositions, or articles. Verbatim notes were classified as written, typed, pasted, or dropped.
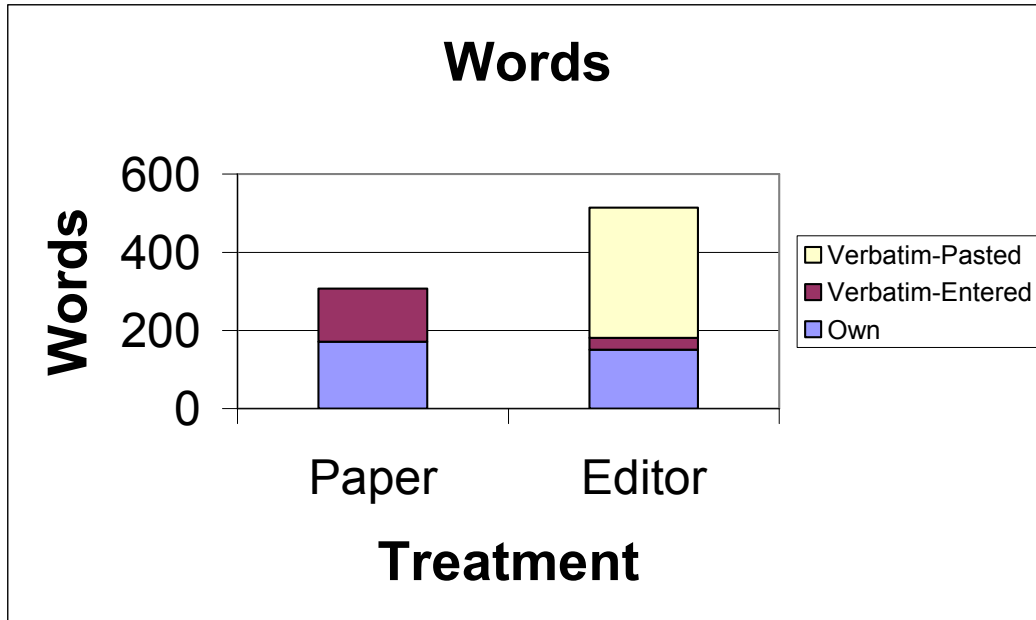
Figure 2: Comparing note quantity and style for text-editors and handwritten notes. The difference in total quantity was significant, as is the difference in total number of verbatim notes. The difference in total number of notes recorded in students' own words is not significant.

Results

Dependent Measures

A repeated measures ANOVA was computed for each of the dependent measures, looking at module by assignment interactions. No significant effects were found for time $F(1, 12) = .0039$, $p>.9$, or number of notes, $F(1, 12) = .036$, $p>.5$. A significant word difference was found, indicating the not-taking tool was used to record more words than pencil-and-paper $F(1, 12) = 7.26$, $p<.05$. No significant effect was seen for the quiz $F(1, 12) = 2.4$, $p>.10$.

Students did not take significantly more notes in their own words on paper than they did using the tool, $F(1,12) = 1.25$, $p>.25$. Tool users took significantly more verbatim notes $F(1, 12) = 10.3$, $p<.01$. There was not a significant difference between verbatim notes taken on paper than those typed with paper $F(1, 12) = 2.4$, $p>.10$. These results are compared in Figure 2.

Survey Results

The survey showed generally positive results for the note-taking tool. Note-taking preference was split evenly between the tool and paper. Nine of 14 students (65 percent) stated that they would use the tool if they were taking the online course, while another said that they might. Two others gave specific circumstances under which they would use the tool- when copious notes were required and when the note-taking tool would allow easier participation in the course.

In free responses regarding their three favorite features of the note-taking tool, half of the students mentioned the ability to paste material from the course materials. Half believed that the tool afforded more attention to the materials at hand, giving responses such as there was "no need to move away from [the] computer" and "one can concentrate on understanding his content [rather] than writing it down."

Another half of the students mentioned the ease of taking notes using the tool. While 6 (43%) felt the tool increased note-taking speed, 4 (29%) felt it was slower. Two students in this latter group claimed to be slower at typing than they were at handwriting.

In free responses regarding their three least favorite features of the note-taking tool 9 of 14 subjects (65%) believed tool use may reduce learning. They reported things such as "typing does not help consolidate my thinking as much as writing", and that the tool "minimized [the] need to summarize ideas because of the 'cut and paste' option."

Three stated that they did not paste often due to this possibility, one stating that he "copied [a] repetitive equation, [but] didn't copy and paste notes [because] I learn them better when I write them down." But 8 of 14 said they made use of the copy-paste feature due to its ease. Two stated that the tool saved time and produced more accurate notes.

Discussion

The hypothesis that the note-taking tool would produce a greater quantity of notes was supported by the evidence that there were more words in notes taken using the tool than there were in those taken on paper. The second hypothesis, that notes taken using the tool would have more verbatim notes than notes taken on paper was also supported. In fact, the increased quantity of notes is entirely a result of the increased amount of verbatim notes, as there was no significant difference in the amount of notes written in students' own words.

The final hypothesis, that the note-taking tool would result in worse performance on module ending quizzes, was not supported. The identical performance of students using the two note-taking styles may also be due to the fact that students did not take significantly less notes in their own words using the tool than they did on paper. Verbatim notes may be especially good for review, as notes taken in ones own words may be less comprehensible upon delayed review than material copied from the course materials. In addition, the digital notes are easier to edit. This allows for easier reorganization of note upon review. The fact that half the students mentioned facilitated editing indicates that they may be likely to use the tool for this purpose.

A majority of students felt that tool use might actually decrease learning. Many of these students related this directly the ease of copying and pasting, one going so far as to recommend that we not allow that functionality. The belief that copy-pasting reduces learning is especially interesting given the apparent lack of difference in actual learning outcomes.

The survey results indicate that this tool would have traction in our courses. Two-thirds of the students said they would use the tool if they were taking the course.

Summary

While previous comparisons of handwritten and typed notes found differences in the quantity of notes recorded, they did not describe the causes of these differences. This pilot study identified copy-paste as a key feature of text-editor based note-taking

applications. Copy-pasting resulted in more notes, of a greater verbatim quality. In general, students liked this feature, to a large degree because it eased the process of taking notes, though many were worried that the functionality would reduce learning. The increase in verbatim notes did not come at the expense of notes taken in their own words or time on task. It may be that copy-pasting is the cause of differences in quantity seen in previous studies.

Previous work both on note-taking and reading in general suggests that the behavioral results of this study could be detrimental to learning. The literature suggests that notes taken in students' own words are superior than notes recorded verbatim, as they reflect elaborative behavior. Though there were no differences on learning outcomes in this study, it may be due to limitations of the design, which was more focused on behavior than learning. The test had few items, all of which were in multiple-choice format. In addition, the test was given immediately following learning activities, and it may be that learning only differs at delayed testing. Finally, the coding of notes with regards to words was somewhat limited. As we have seen above, the wordiness with which ideas are recorded may be a better indicator of note quality. These issues will be explored in greater depth in the following study.

Chapter 3: Experimental Design

Four experimental studies are reported in this thesis. Though each deals with different note-taking interfaces, all follow an identical experimental design. I will describe that design in this section, as well the slight content variations seen in some of the studies.

Procedure

Each study followed a between subjects design in order to compare a number of note-taking interfaces. After informed consent was obtained, students were asked to complete a pretest (in the first experiment, SAT-Math scores were used a substitute). Participants were randomly assigned to one of the note-taking conditions, and given a short description of the tool they will be using. Time was not controlled so participants could take as long as they required to complete the learning material. They were told that they would be able to use their notes to review for a final test on the second day. After completing the materials, students were given the **immediate post-test,** and scheduled to return one week later.

On the second day, students were given a **delayed post-test**. After its completion they were given their notes to study for five minutes. They were required to take all five minutes, and asked to review mentally if they finished reviewing their notes within the five minutes. After studying, they were given a **review test**.

Materials

The learning materials consisted of one module in a course in Causal and Statistical reasoning, which introduces students to concepts of direct causation between variables including definitions of direct causation based on test pairs and response structures. The author (Scheines et. al. 2005) structured the content around ten key ideas and definitions. The module is 15 pages long and contains approximately 9000 words. This is substantially longer than note-taking materials from previous studies

which tend to contain less than 2000 words, and should allow for the evaluation of the coordination hypothesis as concepts can be separated by as many as 14 pages. Appendices A and B contain examples of the course material.

In the first experiment, the content included interactive examples and self-assessment questions. These were removed from later studies in order to increase the focus on note-taking in the context of textual materials.

Testing

Tests consisted of both multiple-choice and free response items. The multiple-choice items required students to use the key ideas to solve problems of causation, such as identifying a variable that directly causes an effect. The free response items asked students to recite definitions and build response structures in which one or more variables are direct or interacting causes of an effect. See Appendix C for examples from the final study. All tests included items of the same structure, though they differed by content. Between studies, item analyses were conducted to identify questions with low discriminability, which were then replaced.

As described above the literature breaks note-taking benefits down into two categories, processing and review. The tests in the studies in this thesis were aimed at these questions. The first two tests were aimed at evaluating the **encoding** effects of note-taking. As students were not allowed to review their notes before either one of them, differences in learning outcomes on the two tests would be ascribed to the different processes of recording notes. While the first test is aimed at immediate learning gains, the second test looks for more robust measures of long-term retention.

The final study evaluated the **external artifact,** or review**,** benefits of note-taking. Students were told at the beginning of the experiment that they would be allowed to review their notes before this test, so could take notes with this goal in mind.

Behavioral variables

A variety of behavioral data were collected. Interactions with the learning material, such as page turns, were recorded and time-stamped. Note-taking behavior was also recorded and time-stamped. Each selection behavior was recorded, including what material was selected, whether the selection was valid with regards to the restrictions, and whether the selection was pasted into notes. Notes were then be coded with regards to which items were recorded. In the initial study, notes were also recorded with regards to wording.

In the first two studies, all coding of notes was done blind to condition by one experimenter. In the latter studies, students could only record verbatim notes. This facilitated automatic coding of note-taking data, which was implemented in an Excel macro. As was the case with hand coding, the macro first split the students' notes into sentences. It then mapped sentences to ideas from the content. Finally, it separated data from key ideas, which were then mapped to the questions in analyses described below.

Surveys

The surveys were aimed at both evaluating students' subjective experience using the note-taking applications and their general beliefs about note-taking. They included both open and closed format questions. Also included were items regarding students' favorite and least favorite features of the tool they used, and whether they would use the tool in a actual course. Students were also asked whether and how the tool affected their note-taking, and whether they thought the tool was beneficial to learning. Questions of preference were asked in accordance to the findings of previous studies, and will be discussed in turn. Examples of survey items can be found in Appendix D.

Chapter 4: Copy-pasting

The pilot study identified copy-pasting as an important function that distinguishes digital note-taking behavior from traditional pencil and paper-based note-taking, and provided data that may explain why previous studies found differences with regards to the quantity of notes recorded. While according to prior literature the increase in verbatim notes produced by copy-pasting behavior would have been expected to be detrimental to learning, the pilot study did not find such learning deficits. Though no differences were found with regards to learning, this may have been due to the limitations of the immediate multiple-choice test or the small sample-size. The study described here was aimed at further investigating the role of copy-pasting in online note-taking. It added the tests described above, and a condition in which students took notes using an embedded text editor which did not allow copy-pasting.

Interfaces

This study was designed to be a more thorough evaluation of unstructured note-taking. Students were assigned to one of three note-taking conditions:

Paper: Students were given their choice of unlined or lined paper to take notes, and were given pencils, pens, and highlighters to use as they desired. The learning material filled the browser in the condition.

Paste: In this condition, students were given a text editor identical to the one seen in figure 1. In this condition students could copy-paste or type notes into their notepad, which took up the bottom third of the browser window. The learning material filled the top two-thirds of the browser. This meant for the text-editing condition, one-third less learning material was visible in the browser at any given time than for the paper condition.

<u>No-Paste</u>: This interface was identical to the one described above except that students were not allowed to copy-paste. They could type, outline, and edit their notes as in the above condition.

Hypotheses

This study was aimed at exploring the impact of copy-paste functionality on note-taking behaviors and learning outcomes. It addressed several hypotheses:

1: <u>If students are prohibited from pasting, they will take fewer verbatim notes and more notes in their own words.</u> In the pilot study, students used the copy-paste functionality to record a large amount of verbatim notes. If this functionality was taken away, they would then be less likely to record notes in a verbatim style and more likely to record notes in the same style as pencil-and-paper.

2: <u>The paste condition will produce shallow encoding of course material, resulting in equivalent immediate performance but reduced performance at a delay.</u> Verbatim note-taking is generally considered to be detrimental to learning. While I did not find this result in my limited pilot study, I believed that verbatim note-taking may indicate shallow encoding, which would be more likely to show up at a delay.

3: <u>The paste condition will result in more accurate notes, increasing performance when students can review their note.</u> Verbatim notes should be beneficial with regards to review, as they are accurate transcriptions of course material. When students create notes in their own words, they are potentially making the information more meaningful, but also introduce the opportunity for inserting incorrect or incomplete information. Reviewing these incomplete or incorrect notes should produce poorer learning outcomes.

Subjects and Materials

A total of 69 subjects from several local universities were recruited by means of a posting to a subject-recruitment website. SAT Math scores were found to account for a significant amount of variability in test scores, and were used in our learning

analyses. Unfortunately, 17 participants did not report their scores, so we were only able to include the data of 52 subjects in our analysis. No students reported being familiar with the course materials, and none employed a "hunt-and-peck" typing strategy. Students were also given a typing speed test before beginning the material.

Each quiz contained 28 items, which tested the 10 ideas on which the instructor based the module and three 12 item multiple choice test which were the basis of our quizzes. Though we matched questions based on ideas, we did not have data to do so statistically. Therefore we completely counterbalanced the presentation of the tests, so that in each condition some would start with test A, others would start with test B, and the rest would start with test C. Two of the tests had 18 multiple-choice items and ten free response items, while the other had 19 multiple choice and 9 free-response questions.

Coding of Notes

The pilot study, this study, and the subsequent study included conditions in which students could record notes in their own words by either typing or handwriting. After notes were transcribed into sentences in excel, they were reordered so that they could be coded blind to condition according to several wording categories. These categories are described below, along with actual examples of the same idea recorded in each style. Wording could then be used in learning analyses of key ideas.

Verbatim: Verbatim notes were of the exact form as in the module.

> *A population has response structure uniformity for a given effect if every individual in the population has the same response structure for that effect.*

Abbreviated: Abbreviated notes had the same words in the same order, but could include abbreviated words or leave out conjunctions, such as "and", or simple prepositions, such as "to."

> *A pop. has resp. struct. unif. for a given effect if every individual in pop. has same resp.*
> *struct for tt effect*

Shortened: Shortened notes had the same words in the same order, but could leave out major words or sections of one to five words.

> *a population has response structure uniformity for a given effect if every individual has the*
> *same response structure for that effect* [NOTE: leaves out "in the population"]

Own: Own notes either used completely different words or word orders. Note that this coding does not require the integration of outside information, which was in fact rarely observed in this experiment.

> *Response Structure Uniformity occurs for a given population if every indivudal has the*
> *same responce structure* [NOTE: spelling errors in notes]

Key ideas were also coded with regards to wordiness, or the number of words the student used to record each idea.

Results

Three main analyses were performed. First, the question of whether the tool had an effect on note-taking was evaluated with regards to both quantity and wording. Analyses also investigated whether there were any differences on test performance. Finally, analyses were performed to link specific aspects of notes to learning
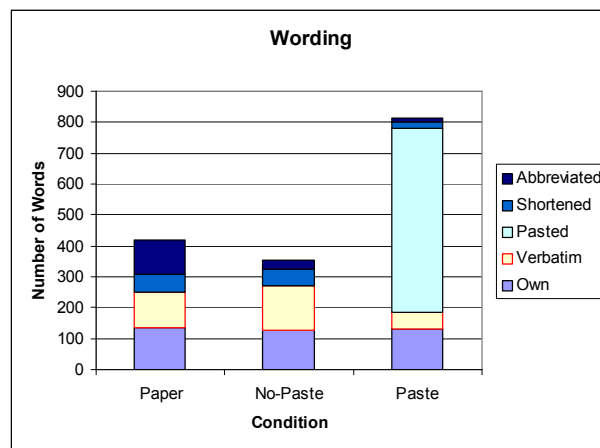

Figure 3: Words by Condition. Students using the Paste tool recorded significantly more words than the other tools. This was entirely caused by the difference in verbatim note-taking.

outcomes. Differences found between conditions cannot be ascribed to time on task, as an ANOVA with condition (Paper, No-Paste, Paste) as the independent variable found no significant difference in module completion time between conditions $F(2, 49) = .024$, $p>.9$.

Notes

ANOVAs were performed on all measures of notes quantity, with condition as the only independent

Figure 4: Total Ideas by Condition. There was a significant difference, with students using the Copy-Paste tool recording more ideas than students using the No-Paste tool. Students in the Copy-Paste condition produced significantly more verbatim ideas than students in the other conditions, while students in the Paper condition produced significantly more own ideas.

variable. Significant differences were found with regards to both note-quantity and wording. With regards to overall note quantity, the Paste condition produced more words than the other two conditions $F(2,49)=7.2$, $p=.001$, (Figure 3), however it only produced significantly more ideas $F(2,49)=3.9$, $p=.02$, than the no-paste condition (Figure 4) The conditions were only marginally different, in the same direction, with regards to the number of key ideas recorded $F(2,49)=2.64$, $p=.08$. The difference between words and ideas is indicative of a brevity difference. Pasters produced by far the most wordy notes $F(2,49)=11.48$, $p<.0001$.

With regards to wording, Copy-Paste was once again characterized by a large amount of verbatim notes. It produced significantly more verbatim words $F(2,49)=13.6$, $p<.0001$, and ideas $F(2,49)=7.93$, $p=.001$, than the other conditions. While there was no significant difference with regards to number of words in the own words category $F(2,49)=.03$, $p=.96$, paper produced more own ideas than the other two conditions $F(2,49)=3.9$, $p=.02$. Though there were some effects of abbreviated and shortened notes, they composed a small enough portion of notes that they could not be analyzed with regards to learning.
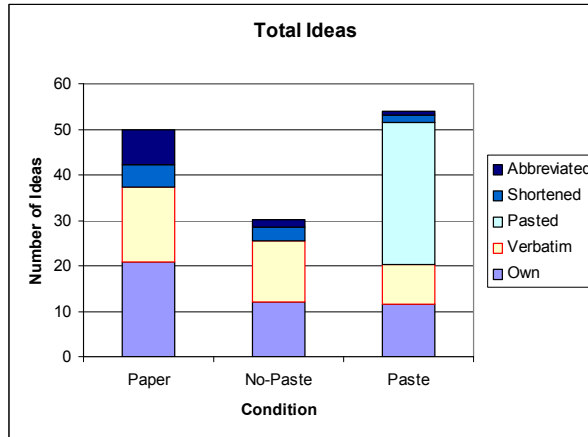
There was a marginal effect of key ideas overall, F(2,49)=2.64, p=.08, with paper producing more than the tool conditions (p=.04 with No-Paste, .07 with Paste). There was a significant difference in the amount of key ideas recorded in students' own words, F(2,49)=4.4, p=.01. However, in contrast with total idea wording, No-Paste did not capture fewer own key ideas than

Figure 5: Free Response scores. The overall interaction is marginal (p=.09), however the forgetting interaction between the first two tests is significant.

Paper (p=.7), though both conditions captured more own key ideas than Paste (both p=.01). The overall Verbatim effect was also significant, F(2,49)=16.9,p<.0001, with Paste producing twice as many as No-Paste (p<.0001), which was not significantly different from Paper (p<.29).

Learning Outcomes

ANOVAs were conducted for all individual tests, including condition, test form and SAT-Math in the model. The latter was included because it was found to account for a large amount of the variability in learning results. No significant effects were found on any of the individual multiple choice or free response tests.

Repeated measures analyses were performed on tests one and two, looking for retention (or process) effects, and on the second and third test, looking for review effects. The overall review effect was marginal for multiple choice tests F(1,34)=2.8, p=.09, and significant for free response tests F(1,34)=15.7, p=.0004 (see Figure 5). However, review appears to be a robust effect, as we found no condition by test interactions.

An overall process effect was not encountered for either test type. However, there was a significant test by condition interaction for free response questions

F(2,34)=4.1, p=.02. Students using the paste condition showed poorer long term retention, forgetting significantly more information than did the other conditions.

Associating Learning with Behavior

We see two major factors distinguishing Copy-Paste from the other conditions that could explain increased forgetting. Pasters took more verbatim notes, and their notes were significantly wordier than those of the other conditions. As described above, verbatim note-taking and wordier notes could help explain reduced retention. As each free response question is tied to a specific key idea, we can treat each question as a single data point, associating it with a specific wording and brevity. We conducted an ANCOVA with Condition, SAT-Mean, Test (including the first two), and Brevity/Wording included as independent effects, and controlling for Subject as a random effect (as each subject would be associated with multiple questions).

There was only enough data to evaluate the "own" and "verbatim" wording categories. We did not find a significant wording by test interactions on the first two tests F(1, 376)=1.4, p=.23, indicating that the wording of key ideas was not responsible for retention loss. Brevity produced somewhat more interesting results. We expected that more wordy notes would be associated with reduced retention, as the pasters had produced wordier notes and forgotten more. However, we found a marginally significant effect in the opposite direction F(1, 876)=3.2, p=.07. It turns out that Paste is the only condition that does not follow this trend. Neither No-Paste nor Paper showed any retention losses between the immediate and delayed tests for wordy items. However, Paste did show these retention losses. It appears that using more words to express ideas did not derive the same encoding benefits for the Paste condition that it did for the other conditions.

Discussion

Hypotheses

Hypothesis one, that typists would record more notes in their own words than pasters, was confirmed. Students without the ability to copy-paste recorded fewer

verbatim notes than did those with the ability to copy-paste. However, while they recorded a higher percentage of key ideas in their own words than the paste condition, this was not true with regards to overall note-taking. Though this may be an indication of focused note-taking, there was no obvious benefit on testing.

There was weak evidence for hypothesis two, that Pasters would perform worse on long term retention due to shallow encoding. While Pasters did not perform significantly worse on the delayed test where they were not allowed to review, they did appear to forget more between the immediate and the delayed test.

There is no evidence for hypothesis three, as all students received significant benefits on the review tests. It appears that students are able to use a variety of note-taking techniques to produce notes from which they can review.

Behavior

It is apparent that the functionality included in note-taking interfaces will affect how students take notes. When given the ability to copy-paste, students created notes of a far greater verbatim quality than when using a text-editor that only allowed typing. This study provides the first evidence that these differences in behaviors may also affect learning. Students using the copy-paste tool appeared to forget more than students in the other conditions, and this forgetting was associated with the wordy nature with which they recorded their notes. Interestingly, though verbatim notes have long been considered negative, the verbatim nature of copy-pasted notes did not appear detrimental to learning, only the wordiness with which ideas were recorded.

Summary

As the different note-taking conditions showed potentially different long-term retention effects, this study may provide some evidence for the encoding benefits of note-taking. This evidence is fairly weak, as only differences in forgetting were observed, not individual tests differences. As all conditions benefited from review,

this study supports the review benefits of note-taking, but students appear to be able to use a variety of interfaces to record notes that are useful for review.

Interestingly, the study did not find any differences with regards to the wording of notes. Students received no benefits from putting notes in their own words. One reason may be the definition of "own words" used in the note-taking literature, which requires only that students reorder the words in the text, or use synonyms. Rewording does not require integration of prior knowledge, which in Kintsch terms would result in the strengthening of the student's situational model. No students were observed to bring outside material into their notes.

Finally, there were effects of wordiness, one of the measures of quantity deemed important in the literature. While wordiness was beneficial when typed or handwritten, it reduced retention when it was achieved through copy-pasting. Wordiness has much smaller cost in the paste tool than it does with typing or handwriting.

The relationship between the value of wordiness and its cost may be related to the focusing hypothesis described in the introduction. The low cost of copy-pasting may allow students to reduce the focus they place on the content they are recording. In fact they do not even have to read the material if they have some other way, such as textual signaling, to identify it as important. The higher cost of typing and handwriting requires students to identify the critical components of the ideas they are trying to record in order to save time. This **feature focusing** may be the foundation of the different retention rates of wordiness in the various tools.

Chapter 5: Intervening on Selection

The work described in the previous chapters found that both wordiness (the number of words used to express an idea) and wording (e.g. verbatim vs. notes in students' own words) were affected by the inclusion of copy-paste functionality in note-taking applications. The literature has linked both of these features of notes with learning, though the evidence is not conclusive for either one. The previous studies found indications of a relationship between the cost of wordiness and learning outcomes. For interfaces where wordiness comes with a time cost, for instance typing or handwriting, wordiness increases learning outcomes, perhaps reflecting increased focus. For the copy-interface, where wordiness has a low cost, retention is reduced for wordier notes, perhaps because wordiness reflects reduced focus. This study explores the possibility of intervening on the selection process in order to increase focus, and thus learning.

The previous study did not find any connection between wording and learning, though that has been posited in the literature. However, when students have the freedom to take notes in their own words or to record them verbatim, they may choose to record difficult or confusing ideas verbatim in order to avoid processing them deeply. The study in this chapter explored whether requiring students to record notes using wording that is different than that in the text will promote learning. Coordinating the two wordings may result in superior learning, and serve the same purpose as recording notes in students' own words.

This experiment also evaluated the effect of intervening on note-taking. As mentioned above, research on traditional note-taking with pencil and paper indicates that only behavioral interventions are effective. Training and instruction are not effective; only when a note-taking style is imposed on a student does note-taking change. The novel interfaces in this study were examples of the control technology offers over the note-taking process.

Interfaces

The following interfaces were developed to evaluate the general research questions described above. In this study, as opposed to the previous study, interfaces in which students took notes using selection-based interactions did not allow typing. This was done in order to increase the focus on the selection interaction. Otherwise, the interfaces would have been susceptible to self-selection effects, where students could choose an interaction method according to the ease or difficulty of the idea being recorded. Students using all tools could markup, outline, and reorganize their notes. In all cases, the notepad took up the bottom third of the browser, and the learning material filled the rest of the browser.

Typing: This condition is identical to the No-Paste condition described above. Students could only enter notes into their notepad by typing. The tool did not give alerts if students attempted to paste or drag contents into their notes; it simply did not let them.

Paste: Students could only copy-paste or drag-drop material from the content into their notes, they could not type.

Restricted: This interface 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.
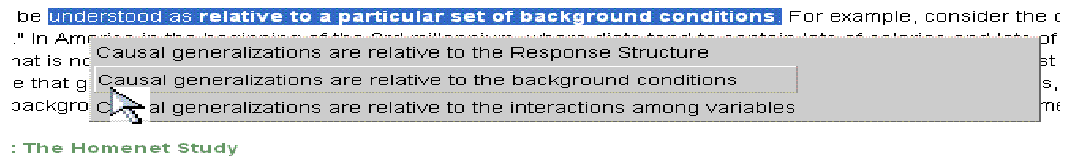
Figure 6: Select Tool. When a student selects text on which they wish to take a note, the interface pops up a box with three options. When one of the options is clicked upon, its text is entered at the end of the students' notes.

Select: In this interface, whenever students made a selection in text, a box with three options became available next to the cursor. Two of the options were distracters, while the third was a reworded representation of the selected concept (see Figure 6). When the student selected one of the entries, it was placed at the end of their notes. 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 such feedback. Few selection errors were observed, however.

Hypotheses

This experiment was designed around two hypotheses:

1: The restricted tool will increase the focus required for students to take notes, and in doing so will improve learning outcomes. The association between large selections and poor learning may be due to reduced feature focusing. Interfaces that require students to focus on the ideas they are recording will increase their encoding of those ideas, thus improving learning.

2: By requiring students to view multiple representations of the ideas they are recording, the select tool will increase performance on learning outcomes relative to tools that allow verbatim note-taking. The idea is that students are required to coordinate two definitions of the same concept, and in doing so encode the principle more deeply.

Subjects and Materials

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 were given incorrect quiz materials.

Figure 7: Time On Task. There was an overall significant effect. Both Paste and Restricted treatments completed the module significantly faster than the Typing condition.

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.

Each quiz contained 25 items, which targeted the 10 ideas around which the instructor based the module and the three 12-item multiple-choice tests that were the basis of our quizzes. In this study, students were also given a pretest, identical in form to the other tests. All other experiments include the same pretests. Notes were coded as they were in the study described above.

Results

ANOVA was performed 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 7). Contrasts showed that Typing took significantly longer than pasting ($p=.04$) and Restricted ($p=.003$) The contrast between Select and Restricted was marginal ($p<.1$).

Learning

In the 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 8, 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 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
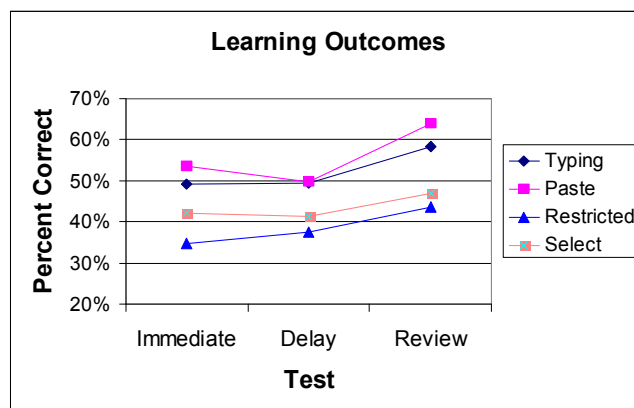
Figure 8: Learning results. There was a significant treatment effect. The two "traditional" tools (Typing and Paste) performed better on learning outcomes than the two novel tools (Restricted and Select).

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.
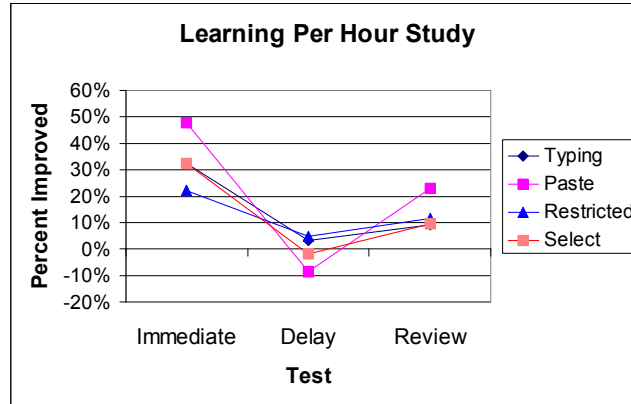
Figure 9: Interfaces differed with regards to learning efficiency, or the time it took to achieve learning results. Performance on the Immediate and Review tests was better for users of the Paste tool.

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 (Figure 9). 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 10), though now the Typing-Restricted contrast is marginal (p=.06). There is an overall wordiness effect as well $F(3,58)=8.0$, $p=.0001$, where Paste is more wordy than all other tools (see Figure 11). 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 significantly fewer key ideas than the non-novel tools (all $p<=.01$), 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 11, where Paste is significantly more wordy than all other tools (all $p<.001$). Typing and Restricted are not significantly different, and though Select is
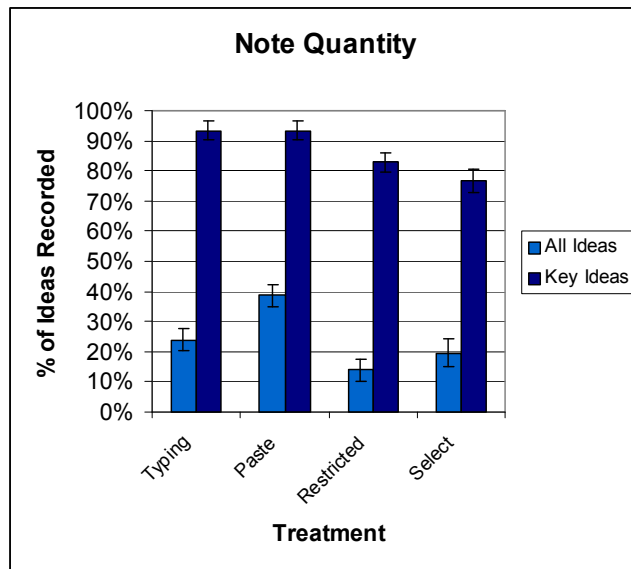
Figure 10: The novel conditions record significantly fewer key ideas, and the Paste tool records significantly more ideas overall. For display purposes, "All Ideas" are divided by the greatest number of ideas recorded, and "Key Ideas" are divided by the total number of key ideas in the module.

significantly less wordy than the two tools (p<.05 in both cases), as mentioned above the Select treatment does not have control over wording.

Associating Learning with Behavior

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

Figure 11: There are significant wordiness effects for both all ideas and key ideas. In both cases, the Paste condition is significantly wordier than the other conditions.

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

Figure 12: There was an overall significant association between presence in notes and performance on learning outcomes. Only the novel tools were significantly better for ideas being recorded in notes.

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 12). 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 interaction $F(1,5362)=7.2$, $p<.01$, the contrasts of which indicate that in-note was only a significant factor 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, with more wordy ideas being associated with significantly poorer performance ($p=.01$, for all other treatment contrasts $p>.5$). This provides support for the hypothesis that the low cost of wordiness for copy-pasters reflects reduced attention to the ideas being recorded.

Survey

The survey included in this study had several measures of students' reactions to the tools. 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

Hypotheses

The restricted interface was designed to encourage students to focus more on the ideas they were recording, with they belief that such feature focusing would increase learning. The selection tool was designed to encourage students to view multiple representations of the ideas they were recording, with the belief that the coordination

of two versions of the same idea would enhance encoding. Neither hypothesis was supported, as both novel tools performed worse on learning outcomes.

Behavior

At a first glance, the restricted interface was a success with respect to behavior, as it reduced the wordiness with which students recorded notes relative to the unrestricted copy-pasting interface. However, overall note-taking was inhibited, as students recorded fewer key ideas when using the restricted and selection tool than either while typing or using the unrestricted copy-paste tool. It appears that as students disliked the tools, the tools were used less.

One of the more interesting results was that eliminating the ability to type in the unrestricted copy-pasting condition increased the efficiency of note-taking, as the copy-paste condition finished the module significantly faster than typists. In addition, the unrestricted copy-paste condition performed equivalently to the typing tool on learning outcomes. While this efficiency is an exciting result, students using the unrestricted copy-paste tool wanted the ability to type. Unfortunately, the previous study indicates that such a combined tool will not achieve the same efficiency gains.

Summary

The study provided further evidence in support of the focusing hypothesis. Students in the unrestricted paste condition did not perform well on items they recorded in a wordy fashion. It is likely that the selection-based interaction technique allows them to record ideas without processing them fully. The efficiency gains observed in the copy-paste tool make the pursuit of effective selection-size restrictions intriguing, but it is clear they must be designed more thoughtfully, as the intervention described in this study failed.

The elaboration hypothesis was not supported by this study, as the two tools that allowed or enforced alternate wordings were not successful. The typing tool was not superior in any learning measure, and the selection tool was obviously less than

useful. The lack of any data supporting the elaboration theory of note-taking led to a concentration on copy-pasting and focusing in the next study.

One of the benefits of technology is the ability to intervene on learning behaviors that previously could not be controlled. The behavioral effectiveness of technological interventions is especially intriguing with regard to note-taking, where as mentioned above only behavioral interventions have been shown to be effective. However, this study provides evidence that such interventions may actually reduce note-taking in general. Students appear to have strong opinions regarding note-taking, and in this case they abandoned interventions they did not like. While the study also indicates interventions can affect learning outcomes, the changes were in an undesirable direction.

This study did not confirm the results of the previous study, which found that copy-pasters forgot more at a week's delay than students using pencil-and-paper or typists. There are several potential explanations for this difference. First, in the previous study students in the copy-paste condition could either type or copy-paste whereas in this study students could only copy-paste. Students may have thus developed a different note-taking strategy in the current study. Secondly, while the previous study had a forgetting effect, this was somewhat weak, as there were no significant differences on individual tests, just between tests. The forgetting effect may have been a statistical anomaly.

Chapter 6: Designing Optional Interventions

The research discussed in this thesis is facilitated by the unprecedented level of control technology provides over the note-taking process. Studies involving pencil-and-paper could not require students to behave in specific ways, as handwriting is by nature freeform. Even if students are given structures (such as matrix notes or outlines) to take notes, outside of the laboratory they can use different paper.

A common fashion of eliciting desirable behaviors in these situations is through pre-training or instructions, which as described above have not been shown to be effective (Kobayashi 2005, 2006). At times, it may be possible to bias particular behaviors, for instance restricting the number of lines students have to record notes in order to reduce the overall number of notes they produce. However, these are not guaranteed to be effective, and are difficult to impose in the real world, where students are free to take notes as they desire. As mentioned above, students are often non-compliant even in the laboratory (Kiewra 1984).

Technology allows us to impose restrictions on how students take notes both in laboratory settings and in the real world. In the research described in the previous chapter, and in the study reported in the next chapter, I investigated the effects of limiting the amount of words students can select in any given note-taking action, and additionally limit the overall number of notes they can record. As these restrictions can be imposed, rather than instructed, inappropriate behaviors can be eliminated. As these technologies can be included in the actual courses, restrictions can be enforced in real learning environments as well as in the laboratory.

However, there are still compliance issues. Students are not required to take notes, much less use the online tool to take them. The integration and affordances already present in the tool, such as the ease of typing or pasting or the ability to access notes

from any computer, may encourage students to use the tool. The restrictions, on the other hands, may actively discourage students from using the interface. Students may simply reduce their behavior, as we saw in our laboratory study where students using the restricted tool recorded fewer notes than students using the standard tools. They may also resort to other note-taking techniques, such as pencil and paper or word processors.

Compliance is not just an issue with optional interfaces. In many circumstances, students can be forced to complete educational tasks. They may be graded, or the interventions may be a part of the curricula. Students can be taken to computer labs, where they are required to complete a set of tasks using a given learning technology. In these cases, educationally effective interventions are guaranteed a high level of compliance. However, tasks can be completed in a variety of ways. Required interfaces may encounter other problems, such as "off task behavior". When students know they have to complete the material, but are frustrated or bored by the intervention, some try to find the easiest way to complete the task. If they know they can easily get an answer by clicking through a set of hints without reading them, they will do so (Baker et. al. 2004).

The problem of adoption in optional interfaces is not 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 (Scheines et. al., 2005). 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 (Hausmann et. al., 2002).

Another study found that students produced superior essays from materials presented two at a time in a split browser than they did when the materials were presented individually (Wiley 2001). However, when the split-frame was not imposed, students automatically readjusted their screen to eliminate the split-frame.

Though the manipulation produced superior performance it appears multiple-frames would not be used by students if not imposed on them.

So while it is clear that students may not always prefer the interface that is most effective with regards to efficiency or learning, when designing elective learning technologies these interfaces cannot be imposed on them. Therefore an interesting design tradeoff exists between desirability from the educational standpoint and desirability from the student standpoint. *The question is how to design elective learning technology that students enjoy using while at the same time guiding them towards beneficial behaviors.* It is not clear how exactly this should be done.

Developers of educational technology have recognized that the traditional methods of design and usability, which tend to focus on immediate performance and user experience, are not always appropriate for educational applications. Educational interfaces are often designed to have "desirable difficulties", which may require more time and produce poor performance on immediate tests, but increase retention and transfer (Bjork 1994). It is recommended that designers of educational technology aim to increase germane cognitive load, while reducing extraneous cognitive load. (Sweller et. al. 1998). In other words, elements that do not encourage learning should be made simpler, while elements that are directly involved in learning should be of more central focus.

Questions of efficiency are especially important for the design of learning interfaces. While traditional user interfaces are generally designed for productivity (correct performance and fewer errors), "an educational activity that learners complete quickly and accurately is of no value if they learn nothing from it." (Lewis 1998) As Gilmore points out, good and efficient performance on an interface optimized for performance may produce low transfer to new or real world problems (Gilmore 1996). One series of studies show how the UI standard of direct manipulation is not always appropriate for educational technology (Golightly 1997). In fact, in these studies it was found that "moving the point of action away from the point of consequence encourages the problem solver to develop their own representation of

the domain", and the difficulty led to more planning on the part of the student. It is important to recognize, however, that these difficult interfaces lead can lead to issues of non-compliance or off-task behavior.

Prior Design Process

The initial attempt to design a restricted elective interface was unsuccessful. It may be that the problem is intractable. However, it is clear that the design process was not optimal. Before describing the process used to design the interfaces included in the remaining studies in this thesis, I will describe how the original tool aimed at restricting selection size was designed, in order to point the procedural flaws made in focusing too early on design candidates too early and collecting limited attitudinal data.

As my previous research had found that lengthy selections were associated with poor learning, it seemed important to dissuade students from making large selections. Restrictions were based on sentences for several reasons. Sentences are easy ways to define individual meaning units, as they have been written as units. It is also easy to automatically parse learning material into sentences, obviating the need for instructors or developers to code ideas by hand.

After having defined a sentence as the basis of the restriction, the basic restriction had to be defined. It could have been as simple as not allowing students to select content in more than one sentence. However, as that could allow students to select entire sentences without focusing on their contents, an alternative was chosen. The initial restriction only allowed students to select ninety-percent of the words in a sentence. It was thought that this would prohibit them from making a selection without carefully processing its contents. In retrospect, the lack of naturalness of this restriction may have simply increased the cognitive load of the selection rather than increasing processing of the learning material.

The next question regarded how to implement the restriction. The two main options considered were "hard-stops", where selections stopped (no more material was

highlighted) when they reached the end of a legal selection. It was felt this would allow students to get around the system by rapidly selecting material until the interface did not allow them to select more. For this reason a different implementation was chosen. Upon the student completing a selection, the system determined whether it was a valid selection. If it was valid, students could copy-paste it into their notes. If not, the selection simply disappeared.

The final step in designing the restricted interface from the previous experiment was conducting a short user test. In this fairly informal test, students used the restricted tool to take notes on the learning materials, and the contents of the notes they produced were evaluated. The tool had the desired behavioral outcome, as the notes were less wordy, and appeared more like typed notes. Unfortunately neither formal survey nor interview data regarding how students felt about their experience with the tool was collected.

The decisions outlined above were rationally motivated, and may have been appropriate if students were forced to take notes in a specific way. However, as note-taking was optional, the difficulty of the interface appears to have led them to take fewer notes. It appears there is a need to tradeoff effectiveness of the behavioral intervention with likelihood of adoption.

Design Requirements

This design process can be improved in several ways. Firstly, **design decisions should be more closely attached to user data**. The data I have regarding effective self-restrictions should form the basis of the restrictions I design into my application. While there are several attractive reasons for using sentences as the basis of restrictions, is important to validate this with data from students with effective strategies. Student data may also help in deciding whether selecting entire sentences is in fact an appropriate behavior, which would allow us to avoid unintuitive restrictions at the sub-sentence level.

In addition, **the user study must evaluate multiple interfaces and collect experiential data.** Evaluating multiple interfaces will allow me to understand the tradeoffs between different interface features. Once again, student data should drive the selection of a proper interface. Students' experience using the tool must also be evaluated. It will do no good to produce an interface that students will not use. In the end, we may have to select an interface that produces suboptimal behavioral results in order to increase the likelihood that it will be used.

Design Process Overview

The design process described here was initially developed to solve the specific design problem encountered in the previous study: how to restrict copy-paste selection size in a way most likely to be adopted by students. However, the process is also intended to be used for more general design problems where the goal is not restriction, but simply to support a specific behavior.

Here I give a general overview of the design process, which follows an iterative procedure (Nielsen 1993) and makes use of the think-aloud protocol (Boren & Ramey 2000), both of which are common in the human-computer interaction literature. The overall steps were identical for both studies, but the process was significantly refined between studies. Details are filled in for each study. Note that this is not intended to be an exhaustively empirical evaluation of all possibilities. Instead, it is intended to rapidly explore a design space, testing hypotheses about the design space through iterative design.

Step 1: Describing the Design Space

The first step is to define the **design space** by a set of **dimensions** that can be varied to produce different interfaces. For example, restricting selection size can vary with regards to the *timing* dimension; a selection can be imposed while a user is making a selection, or after a student has finished making a selection. They can also vary with regards to whether the restriction is *explicit* (i.e. a popup saying that an inappropriate selection is not allowed) or *implicit* (i.e. changing an inappropriate selection without giving the user written feedback). As these two dimensions are

both binary, a design space that only varies these dimensions would have four possible interfaces.

In order to avoid missing out on design opportunities, these dimensions should describe the design space as thoroughly as possible. In the initial design study described here, the dimensions were identified mostly by developing as wide a range of interfaces as possible, and then describing the dimensions on which they differed. In the second study, a large set of interfaces already existed. By comparing their functionality, I was able to determine a set of interface dimensions that described them.

Step 2: Developing Test Interfaces

Once the interface dimensions have been defined, a set of interfaces is built that combine different levels of the dimensions in different ways. Enough interfaces should be built so that every level can be evaluated in more than one interface. Multiple combinations are intended to avoid potentially misleading interactions between levels of dimensions, where only the combination of two dimensions leads to a negative result.

Step 3: User-Testing Interfaces

The test interfaces are then be evaluated in think-aloud user studies, with each participant using a set of tools that will allow them to compare multiple levels of the same dimensions. Interface errors, expressed frustrations, are recorded and associated with their respective interface dimensions. The think-aloud protocol is a standard observational technique for usability derived from cognitive science (Boren & Ramey, 2000)

After using the tools, the users are asked to complete a survey containing both closed and open-ended questions regarding their experience using the tools. They are asked to rank the tools with regards to their desirability, and asked which features they liked the most and least for each tool. Participants are also asked whether they

would use the tool in an online course, how the tools affected the participants' note-taking processes, and whether they thought the tools were useful to their note-taking.

Step 4+: Iterate

The goal of the user tests is not to reach statistical significance on any rating, but to collect enough data to make reasonable judgments regarding the behavioral impact and user satisfaction of specific interface dimensions. Once these judgments are made new interfaces are then developed to validate those judgments. This can be considered a sort of hypothesis testing. These interfaces are then submitted to further user testing, identical in form to that in step 3. This process continues until the designer is satisfied with user responses. Such iteration is a basic and effective principle of interface design (Nielsen 1993).

The goal of iteration is to test **design hypotheses**. In the following studies, I will describe situations in which early testing revealed contrasts between stated preference and observed behavior, occasionally occurring when users rated unfamiliar dimensions (dimensions for which they had only experienced one value) in a way that contrasted with the experience of users who had experience using both dimensions. The hypothesis that experience with the dimension would cause users to rate the dimension differently. Subsequent iterations would test this hypothesis by giving all users experience with the interface. The goal is not to achieve statistical significance, but to gather further user data on areas of interest from previous iterations.

User Goals

An important feature of this study is the evaluation of users' goals and motivations. The think-aloud is intended in part to get readers to express the reasons they have for note-taking behaviors. The interviews and questionnaires included in these studies are also designed to elicit information regarding these topics.

Educational Design Studies

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" (Collins et. al., 2004). However, these design studies are large scale classroom interventions, whereas this process was initially focused on rapidly developing optional interfaces for experimentation. It is aimed not at replacing large-scale classroom studies, but at helping ensure that interfaces included in classroom or laboratory studies are grounded in user data as well as theory.

Chapter 7: Redesigning Restrictions

The restricted interface evaluated in chapter 5 was designed to encourage shorter selections in copy-paste note-taking, which the previous two experiments in this dissertation have shown are superior to large selections with regards to learning. This interface inhibited note-taking behavior, as its users recorded fewer notes. This chapter describes the redesign of the restricted interface. In order to develop a restrictive interface more likely to be adopted than the one reported in the previous study, it was necessary to explore the tradeoff between the behavioral effectiveness of the intervention and user satisfaction.

I used the method outlined in the previous chapter to redesign the interface. The work was intended to evaluate the impact of different interface dimensions and study how students interacted with different note-taking interfaces. It was also intended to assess the motivations behind students' behaviors through observation and interviews. The best interfaces, which showed highest user satisfaction while still reducing selection-size, were included in the experimental evaluation described in chapter 8.

The first concession made to user satisfaction was to base all restrictions on the sentence level, whereas the previous design limited students to a percentage of a sentence. As sentences are easily recognizable units, restrictions based on them should be less confusing than restrictions based on a percentage of words in a sentence. The first stage of our study included the initial restricted interface from the previous study for comparison purposes in order to confirm this decision.

Subjects and Materials

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 research university.

Participants were tested using Internet Explorer on a Windows XP desktop with a 17-inch monitor. In our descriptions of the data, I identify relevant subjects by using their subject ID (i.e. Subject 1 = S1).

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.

Interface Dimensions

In this first iteration of the design process described in the previous chapter the interface dimensions were identified by abstracting from the designs developed for the first user-tests. I will describe those dimensions here:

Recommend/Restrict: This dimension distinguishes between interfaces that *recommend* a certain behavior, and interfaces that are *restrictive*, or which in other words enforce the behavior. 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.

In/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.

Single/Multiple Sentence Restriction: The third dimension 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.

Implicit/Explicit 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.

First Round Interfaces

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

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.

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 (see Figure 13).
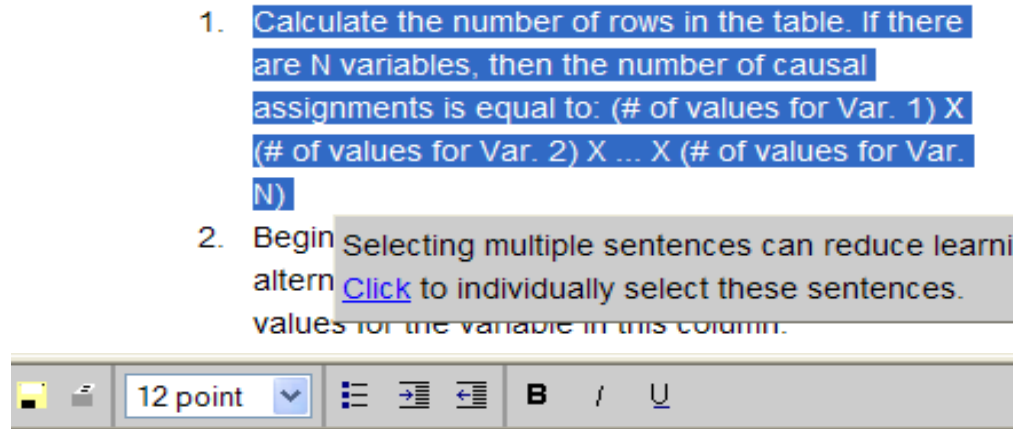
A response structure must contain **all** possible causal assignments

Figure 13: Example of a selection-based interface recommending that the user select fewer sentences. When the user makes a selection that spans multiple sentences, the interface pops up this recommendation. This shows the *recommend* level of the interface. A *restrict* level would not allow a multiple sentence restriction. With regards to timing, the interface pictured here pops up the message *after* a selection is made. If the interface acted *during* the selection, it would popup the message as soon as the user crossed the sentence boundary. Finally this is a *multiple* sentence interface, as clicking the link reselects all the sentences individually. A *single* sentence interface would reselect the first sentence if the link was clicked.

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.

Original: 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.

Initial Testing Results

One of the more unexpected results from the 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 the 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). This behavior, which I call **selection-to-read**, played a large part in some of the conclusions derived from these initial user-tests, which we will now describe.

It appears to be 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!" (S1). This led me 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 text they had originally selected 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.

Changing the users' 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 this 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 it was no students' least favorite interface. 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.

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. We were interested in evaluating three design questions:

1. How did users respond to an interface that combines the features of the click-sentence and hard-stop interfaces, two of the more popular interfaces

	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 of restrictive interfaces

from above? This interface was discussed in the interview with subject 3, who had encountered click-sentence functionality in the hard-stop interface.

2. Can we design a multiple-sentence reselection that does not result in transcription errors?

3. How can we reduce the number of recommendations that appear for non-note-taking behaviors?

The interfaces tested in the second round are described below:

Click-and-Select: This 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 the selection of 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.

Revised the Reselect-Multiple: Changing the user's selection violated a guideline from the first round. However, the multiple-sentence reselection only adds unselected spaces between sentences. This appeared to be less frustrating than the single-sentence reselection because it allowed selection-to-read. Unfortunately, the initial interface produced transcription errors. We hoped to reduce these 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.

Revised Recommendation: In this tool warnings were only displayed post-process, in order to avoid popups caused by selection errors. Initially, the interface only displays warnings when users copy or begin to drag an inappropriate selection. Doing so 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. The calculation 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. I was hopeful that this carrot-and-stick approach would encourage shorter selections.

Candidate User Tests

The click-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. This provides validation for the guideline against modifying users' selections.

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.

Summary

This study had several positive outcomes. It produced a series of recommendations regarding how best to intervene on selection, consolidated in Table 1. While may of them may be intuitive to a user interface designer, it is important to remember that interface design for educational technology occasionally requires that interfaces be somewhat difficult as long as the difficulty increases "germane cognitive load" (Sweller et. al. 1998, see chapter 6 of this document for more detail). However, the guidelines developed here were generalized from situations in which inappropriate difficulty was encountered, for example when an interface caused a user to make a transcription error, or when the intervention caused problems for unintentional behaviors. As these difficulties are not learning events, they should be avoided.

By observing behavior, the study also allowed for the detection of a behavior that has not been described in the literature. Students were observed to select text in order to help them read, and reported doing so in order to increase attention.

The click-and-select and final recommendation tool appear to be effective both with regards to behavior and user satisfaction. The click-and-select tool should be included as a popular restrictive interface. The modified recommendation tool also achieved a high level of user-satisfaction; however it is unclear whether the recommendations would be enough to modify behavior.

Chapter 8: Restricting Selection

The experiment described here is intended to evaluate whether restricting the amount of text students can select in a single copy-paste action promotes learning. This was also one of the goals of the experiment described in chapter 5. In that chapter, the restrictive interface suffered poor adoption due to bad design. It was hoped that the design process described above would increase adoption. A second goal of the experiment was to determine whether recommendations were effective in reducing selection size.

The click-select tool was the preferred restricted interface in the design study, so it was included in this evaluation. 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 the tools were designed to promote single-sentence selections.

Interfaces

The following interfaces are included in this study:

Click-Select: Students were only allowed to record notes in their notepad via copy-paste or drag-drop. They were restricted with regards to how much they selected. Once students started a selection, they could not select beyond the boundaries of the initial sentence. They could also select an entire sentence by simply clicking on it once.

Recommendation: This interface popped up a warning every time the user copied a selection containing multiple sentences. If that sentence was pasted, the interface then popped up a warning for every inappropriate selection. This behavior disappeared as soon as the student performed another short copy-paste action. Students were not given a link to reselect the sentence, as it had not been clicked in the user tests.

Unrestricted: In this interface, students could copy-paste any material.

None: In this condition students were asked to read the material without taking notes. Students in the no-notes condition were asked to mentally review between the delayed and review tests, a standard control procedure in note-taking studies (e.g. Carter & van Matre 1975; Fisher & Harris 1974; Rickards & Friedman 1978).

Hypotheses and Goals

This experiment was designed to evaluate several hypotheses:

1: 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.

2: 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.

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

4: The novel interfaces will increase the focus required for students to take notes, and in doing so will improve learning outcomes. I believe that the association between large selections and poor learning is due to reduced feature focusing.

Interfaces that require students to focus on the ideas they are recording will increase their encoding of those ideas, thus improving learning.

This study was also aimed at using a larger subject pool to validate the findings from the design study. Of key additional 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.

Subjects and Materials

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.

Each quiz contained 21 items, which targeted the 10 ideas around which the instructor based the module and the three 12-item multiple-choice tests that 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.

Results

Behavior

ANOVAs were conducted on each note-taking measure, with condition as the only independent measure. Pretest was not included in the model, as it was not found to be a significant covariate. Time on

Figure 14: Note Quantity. The unrestricted condition recorded significantly more notes, with respect to both words and ideas, than the other two note-taking conditions.

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

Figure 15: This graph shows the percentage of ideas that were present by themselves in notes. The unrestricted condition had significantly fewer single-sentence selections.

by each tool (see Figure 14). In both cases, contrasts showed that the unrestricted tool recorded a greater quantity than the other tools, which were not significantly different from each other. There was not a significant difference with regards to the wordiness with which users of each tool recorded their ideas $F_{(2,37)}=2.07$, p=.14. However, there was a significant difference in whether users of different tools copy-pasted individual ideas or groups of ideas in one selection $F_{(2,37)}=8.17$, p=.001. Once again, contrasts indicate users of the unrestricted tool were significantly more likely to record ideas together, rather than selecting each one individually (see Figure 15).

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.

Figure 16: Learning Results. There were no significant differences on individual tests. The click and unrestricted tool were the only interfaces to benefit significantly from review.

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 $F(2,2347)=7.16$, $p<.001$. Contrasts investigating the treatment interaction indicated that presence was only a significant factor for the two novel tools.

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.

Figure 17: Interface ratings. Click-select was significantly higher for accomplishing goals and experience. When averaged together, the ratings are also significantly different.

The click select tool appears to be better received than the other interfaces (see Figure 16). There was a significant difference for accomplishment, χ^2 (2, 38)=6.9, p=.03, a marginal difference for Experience χ^2 (2, 38)=5.2, p=.07, and no difference for learning χ^2 (2, 38)=2.23, p>.3. Averaging scores across all measures is also significant χ^2 (2, 38)=9.09, p=.01.

The survey asked our fifty-three 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.

Discussion

Hypotheses

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, which predicted higher satisfaction for the recommendation tool but better compliance for the restricted tool. 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 using 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, reduced note-taking may be beneficial over an extended period of time, 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 the other tools. 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. The preference for the click-select tool 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. Instead of attempting to encourage behavior through interface design, designers should ensure that the interface allows for the simple collection of behavioral data that is associated with learning outcomes. This data can then be passed on to additional systems, which could use the data to give additional instruction or assessments.

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.

Summary

Recommendations from Design Study

I 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. It is interesting that recommendations are enough to modify user behavior without reducing user satisfaction relative to an unrestricted interface. When done unobtrusively, nagging users can encourage desirable behaviors. However, our nag interface was still not as satisfying as our restrictive interface.

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 gave 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 (Marshall & Bly, 2005). 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.

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 (Corbett et. al. 1997). 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.

Chapter 9: Highlighting and Selection

The results of the experiments described up until this point have focused on copy-paste based note-taking, where students take notes in a separate document. Two main findings have led to this focus. First, I found that copy-paste functionality has the potential to increase the efficiency of the note-taking process relative to typing, so designing interfaces that encourage that behavior may save students time. Second, there appear to be clear links between how students copy-paste notes and their performance on learning outcomes. Specifically, they are more likely to remember key ideas they record, and they are more likely to perform better on test items covering key ideas they record in smaller selections than key ideas they record using larger selections.

The results of the previous experiments indicate that intervening to change how students can select material to copy-paste is an inappropriate path. First, most students want the ability to type in their notes even though it does not appear to help them learn, and is slower than copy-pasting alone. It is unclear whether students would use a tool that only allowed them to copy-paste, as they rarely use even an unrestricted tool in courses taken online for credit. Secondly, even when we can promote desirable behaviors, it has no impact on learning, though it can increase user satisfaction. As mentioned in previous chapters, it may be better to use student behaviors as a window into student knowledge, rather than intervene to produce behaviors that are positively associated with learning.

The remainder of this thesis compares highlighting with copy-pasting. As both use selection-based interactions to record notes, their behavioral and learning outcomes may also be related. Because highlighting does not traditionally involve typing or handwriting, students given highlighting interfaces may not miss the ability to type. If highlighting performs similarly to copy-pasting, it may be able to increase efficiency in realistic note-taking applications. As I will detail below, the main difference between highlighting and copy-paste based note-taking, the notepad, also allows for

the further investigation of the cognitive mechanisms underlying the positive benefits of note-taking.

Note-Taking, Copy-Pasting, and Highlighting

Using text-editors to record notes is similar to handwriting notes on paper, as both involve creating a personalized set of notes independent from the learning materials. This can be referred to as **unanchored** note-taking. Copy-pasting behavior has no paper-based analogue in traditional unanchored note-taking. However, when students annotate documents they behave in ways similar to copy-pasting. This is **anchored** note-taking, or note-taking in which students take notes on the document they are reading rather than create a separate document. This language is taken from work on discussion boards, where anchored discussions are integrated with textual content and based on individual highlights, and unanchored discussions occur on separate pages from the content (Brush et. al. 2002).

The basic copy-paste interaction involves selecting content with the mouse, and then dragging it into a notepad. This is analogous to highlighting and underlining. In these annotation techniques students use a tool to select material, which then becomes visually distinct from the rest of the content. Digitally, the initial process of selecting text is identical for most highlighting and copy-pasting interactions. In other words, students use the mouse to select text, which is then either highlighted or copy-pasted using some basic note-taking interface. Examples of these interfaces will be explored below.

As highlighting and underlining appear to have the same behavior and learning outcomes (Fowler 1974), I will refer to both from here out as *highlighting*. Reviews of note-taking, especially in the educational domain, often treat highlighting and note-taking with paper (which I will refer to as *note-taking*) as part of a larger group, similar but distinct (see Wright 1988; Caverly 1991; Anderson & Armbruster 1984). Articles report experiments comparing the two (Lonka 1994, Annis & Davis 1978), or individual studies reporting each behavior distinctly (Ooostendorp). Some studies indicate that highlighting is faster (McAndrew 1983; Kulhavy 1975), and while it

performs worse on constructed response items, it appears to increase performance on multiple-choice items (Kulhavy 1975). While the studies mentioned above provide some evaluation of the different behavioral and learning of note-taking and highlighting, no studies look at whether and how the students perceive the behaviors as different.

Copy-pasting is one type of digital note-taking behavior, but one that the results cited above may be an appropriate behavior to encourage, as it produces similar learning outcomes in less time. If highlighting is similar to copy-paste, the question remains as to whether the results described in the copy-paste experiments above will transfer to highlighting interfaces. If this were the case, highlighters would be expected to complete modules as quickly as copy-pasters. In addition, they would be more likely to retain ideas they recorded than ideas they did not record, and perform better on shorter highlights than larger highlights.

The obvious difference between highlighting and copy-pasting is the existence of a notepad. This is also a major difference between traditional highlighting and note-taking. It may be that the presence of this notepad plays a role in the positive learning outcomes achieved by note-taking. A notepad may allow students to easily **coordinate what they are currently learning with what they have already studied**. In other words, when students are reading a novel concept, they can easily refer back to previous material by looking at their notes, rather than flipping back pages. The easy access may encourage such coordination. Such simultaneous availability of multiple learning sources has previously been shown to improve learning outcomes such as essay writing (Wiley 2001). It may be that easy access to multiple related documents allows **bridging inferences** to be made by students. Evaluating the value of a notepad will address these questions regarding the elaboration component of note-taking.

Highlighting Studies

The research described in the following chapters addresses the following three questions:

- How are copy-paste based note-taking and highlighting similar with regards to behavior and performance on learning outcomes?
- How does the presence of a notepad affect note-taking behavior and learning outcomes?
- How do students perceive the relationship between note-taking and learning?

The first step in addressing these questions involved developing a highlighting interface to be included in an experimental comparison with a copy-paste application. I developed the highlighting interface using a refined version of the design study described above. This design study will be described in the following chapter. I will then report the results of an experimental evaluation that addresses the above questions by comparing highlighting, copy-paste, and an interface that integrates highlighting and notepad functionality in order to assess the impact of the notepad by comparison with the highlighting only interface.

Chapter 10: Designing Highlighting

The design study described here had three major goals. First, it explored the design space available to developers of online highlighting tools, and gave direction regarding the impact of specific interface dimensions on which highlighting interfaces can be varied. Secondly, it explored why students highlight, and what their goals are when they highlight materials. Finally it was intended to produce an easy-to-use highlighting interface to be included in an experiment comparing highlighting with copy-paste note-taking.

The process described here is a refinement of the one used in the previous study. New questionnaire instruments were added. In addition, it deals with a distinct type of design problem. The previous study was intended to produce a user-friendly restrictive interface. In this study, there was no intention to restrict the user. Its goal was simply to design the most user-friendly highlighting interface possible within the design space. In the latter case user response takes priority, whereas in the former case the design process is intended to tradeoff behavioral effectiveness, or whether the interface effectively produces the intended behavior, with user satisfaction.

Describing the Design Space

The highlighting interface was designed for a standard desktop computing environment. Users of the highlighting interface interacted with standard browsers using a keyboard, mouse, and computer monitor. The mouse was used to create selections that determined the highlight to be created. This space is distinct from a design space that includes stylus-based interaction, such as tablets (e.g. Schilit et. al. 1998) or PDAs (e.g. Davis et. al. 1998).

There is, of course, a real world equivalent to highlighting. Readers often annotate text with a "yellow highlighter." There are several major differences, however. Highlighters are not by their nature dual-use tools. While some students may use them to write words, the tools' primary task is to highlight. In contrast, the mouse

has multiple functions, and highlighting may not take priority. The mouse is used to navigate and make menu or button selections that have nothing to do with text. Highlighting is an additional feature of the mouse, whereas it is the primary function of a physical highlighter. Note that these multiple responsibilities also exist for most styli, which are used for navigation and menu selection as well as annotation. It is true that a pencil can be used for both underlining and writing text, but both tasks are arguably note-taking, and definitely informational. A mouse has both informational and navigational responsibilities.

A second major difference between digital and traditional highlighting regards the ability to manipulate highlighted text. On paper, a highlight tends to be permanent, whereas there is no reason a digital highlight cannot be deleted or extracted. The ability to extract highlights into a separate notepad means that the highlighted content can then be reorganized or edited.

Process

The first step in the design process (outlined in greater detail in chapter 6) involves identifying a set of interface dimensions that describe the design space. The high-level description of the parameters within which the highlighting tool was designed outlined above tells us nothing about how a highlighting tool can be designed. In order to be useful, the space must be described with regards to interface dimensions that can be manipulated independently. In the previous study, these dimensions were determined by describing a set of potential designs, and abstracting the dimensions on which the designs varied. This study represents an attempt to be more systematic about defining the appropriate interface dimensions. Fortunately, a variety of web annotation tools have been built for both research and commercial purposes. Studying the similarities and differences between these tools allows for the creation of a set of dimensions that describe them.

The second step of the design process was the creation of a set of interfaces that cover different combinations of the different dimensions. Creating all combinations would be intractable beyond 3 dimensions, so the goal was to produce enough

interfaces to cover all levels of the dimensions in multiple interfaces. I will describe below how I chose which potential interfaces to implement. It is important to note that these interfaces should not be thought of as final. Instead, they are used to explore users' responses to different dimensions and their combinations.

These interfaces were then submitted to user test using the think-aloud protocol, the third step in the design process. As in the last study students did not appear to verbalize much of their note-taking activities. In this study I included instructions to read-aloud, whereas in the previous study students were explicitly told they were not required to do so. Students used three tools in sequence, and tools were be assigned so that users would get experience with as many levels and combinations of the different dimensions as possible. After using each tool, students were asked to rate the tool and describe their favorite and least favorite features. Appendix I describes both the coverage of dimensions for the interfaces and the assignment of interfaces to participants.

After finishing the final tool, users were be asked to fill out a questionnaire that asked for their opinions regarding each variable on a 7-point Likert scale. For example the question "Do you prefer to initiate the highlight action (e.g. click a button) before or after you make a selection?" had answers 1 ("strongly before") to 7 ("strongly after"), with 4 being neutral. Where dimensions had multiple levels, users were asked to pick their preferred level, and then rate the importance on a 4-point Likert scale. The survey is presented in Appendix G of this document.

The final portion of the user-test involved a semi-structured interview, which covered several topics. Main areas of interest included the goals students have when highlighting material, what type of material they are likely to highlight, and how they use their notes. The conversations started with a general question of whether students highlight, and moved on to the goals they have if they do in fact highlight. This process often led students to talk about what they highlight, which the interviewer used to prompt further exploration. For example, the interviewer may ask a question: "You mentioned you tend to highlight definitions and key words. Can

you think of anything else you highlight?" If the interview did not touch on a topic of interest, it was raised directly by the interviewer with prewritten question forms.

The data from the user tests was analyzed in several ways. During the user test, the evaluator made note of difficulties users had with the interfaces, as well as verbal comments regarding either the interfaces or the users' highlighting strategies. These were then confirmed and transcribed from the screen and audio capture. Major interface problems were fixed within the first phase, and are described below. Interface dimension ratings were collected and averaged. Three interfaces were then created to test theories regarding behavior observed in the first phase, and ensure full coverage of the interface dimensions. These were then submitted to user testing of identical form to the first user-tests in this study.

Subjects and Materials

Six subjects participated in first phase user-tests (S1-S6), and 5 participated in the second phase (S7-S11). 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, I identify relevant subjects by using their subject ID (i.e. Subject 1 = S1). Each participant completed a 15-page module in Causal and Statistical Reasoning using three different interfaces to record their notes.

Step 1: Exploring the Design Space

Though a range of highlighting interfaces have been built for commercial and research purposes, there are no guidelines regarding how highlighting should be implemented, though there are a variety of ways highlighting can be supported. Reviews of the literature often address more general questions of annotation such as sharing (Wolfe & Neuwirth 2001). Ovsiannikov reviewed 17 commercial and academic annotation tools and produced up with a range of architectural, functional, and user-interface recommendations (1999). Ten of these applications offered highlighting functionality, as did the one he developed and tested with users. Still,

there are no guidelines regarding how to create a user-friendly highlighting interaction. As will become clear, there are a range of methods for supporting highlighting, but no clear evidence regarding best practices.

In order to determine the different ways in which the highlighting interaction could be supported, I evaluated a total of 22 computer-based annotation tools that support highlighting (see Appendix H for a full listing). I used any tool I could find that either adequately described the interaction technique in research or instruction materials, or which could be used. Ten of these are commercial applications, all of which are focused on web-based annotation. Twelve are applications built for research purposes. Three of these were specifically built for web-based education, 2 of them for general purposes (including non-web documents), and 7 of which were web-based general-purpose applications. Most tools included highlighting as a central technique, but included additional functionality such as attaching typed comments to highlights. Many research tools had to be eliminated for not adequately describing the highlighting interaction.

Interface dimensions are determined by comparing the different interfaces. For example, in order to create a highlight in YAWAS a user first makes a selection, and then right clicks on the selection to get a context menu, then clicks on the highlight entry in the context menu (Denoue & Vignollet 2000). To create a highlight in Annotizer, the user first makes a selection, and then clicks a button that is on a toolbar to the side of the text. Here the tools differ with regards to the visibility of the highlighting action. In the case of Annotizer, the button is present as a visible reminder on the screen, whereas with YAWAS the context menu must be opened by the user. Both can be compared with Diane, where to create a highlight students first click a button on the top of the page and then select text, which is automatically highlighted (Bessler et. al. 1997). This differs from both Annotizer and YAWAS in that the selection is made after clicking the button rather than before. It differs from Annotizer in that the button is at the top, rather than the side.

One highlighting interaction used in several applications involved creating predetermined locations that allowed highlighting. These points could be a button, sentence, or predefined locations embedded in the text. By clicking these buttons, the word, sentence or arbitrary amount of text was highlighted. I did not include this dimension for several reasons. First, it does not allow for the freedom seen in traditional highlighting. Second, it can require the content creator to identify appropriate points for annotation, which reduces the circumstances in which the interface developed in this study could be deployed.

Interface Dimensions

Comparing the interfaces as described above resulted in the identification of five dimensions that could be manipulated:

Action Timing: The highlighting action can take place either before or after a selection is made. In the former case the user would select text with their mouse and then take an action that causes the selected text to be highlighted. In the latter case the user would then take an action such as clicking a button, and the next text selection they create will automatically be highlighted.

Visibility of Action Source: The user must take some keyboard or mouse action in order to initiate the highlight action. A visible action source is involves clicking on something that is visible on the screen, most often a toolbar. Note, however, that a toolbar can become invisible if the user scrolls and the button does not follow. An invisible action source requires the user to take an action that is not apparent on the screen. In this study we include clicking on the *context-menu*, *keystroke* (pressing an arbitrary key), or *left-clicking* the selection.

Action Duration: Once a highlight action is initiated, it can either be active for a single selection or for all subsequent actions until highlighting is turned off. This is analogous to picking up a highlighter in order to annotate documents; the highlighter can be used until the reader sets it down.
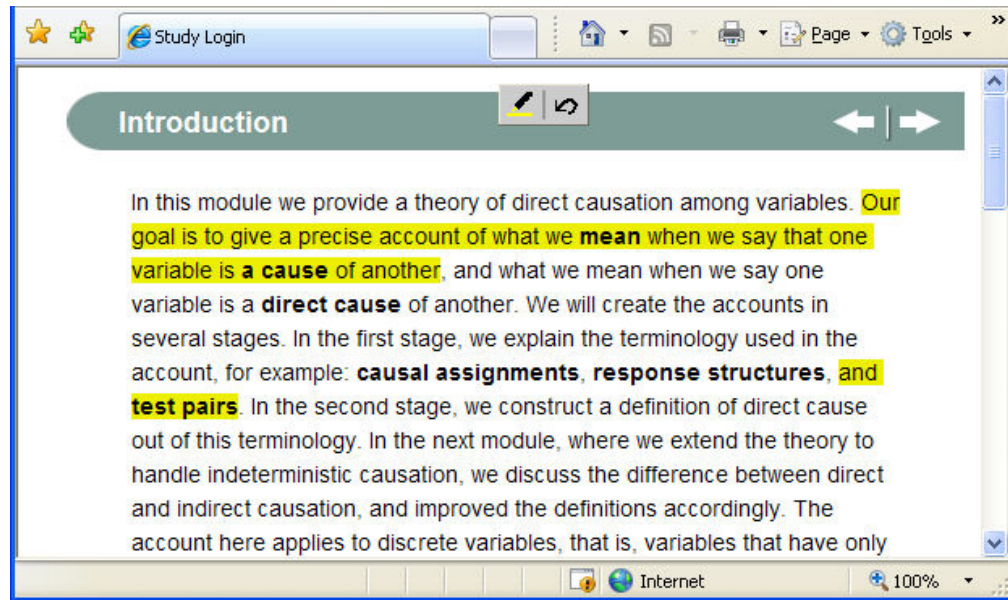
Figure 18: Press Button Before interface. In order to make a highlight using this interface, students first press the highlight button, then make a selection. If this had an *after* value for the timing dimension, the user would first make a selection and then press a button on the top, which is the Press Button After interface from the intial user tests. This has a value of *single* for the duration dimension, as it only acts on a single selection. A *multiple* value for duration would cause every subsequent selection to become a highlight until the user clicks on the highlight button again. This interface has a *permanent* value for the contextual/permanent dimension, as the menu-bar was always available. If it were a *contextual* interface, it would only be available when a selection is made.

The following dimensions only apply to visible action sources like buttons, not invisible actions like key presses:

Contextual vs. Permanent Action Source: A permanent action source is always available on the screen. A contextual action source is only present on the screen when the user can make an action. In the case of highlighting, a contextual menu would appear when the user selects text with the mouse (in "After" timing), and can then use the menu to highlight the selection. When no text is selected, the highlighting toolbar is not visible.

Location of Action Source: The note-taking interfaces evaluated here have toolbars in three locations, either the top of the browser window, on the side of the browser window, or right next to the mouse.

Initial Interfaces

There are over 80 possible combinations of different levels of the above dimensions. Implementing and testing all of them would be very time consuming, and defeat the purpose of a rapid design study. It is important to reduce the total number of interfaces to test, while still providing coverage of different combinations of all dimensions.

Fortunately, there are a variety of combinations that do not make sense. First, contextual actions are not possible when the action must be taken is taken before selection, because by definition a contextual action is only available after the selection is made. Similarly, the context-menu should not be permanent, because in standard operating systems it is a contextual event. Permanent actions should not occur after the action, because it is not associated with single actions, but rather a series of actions. Another argument can be made that permanent actions should not involve invisible sources, because the user would have to rely on memory to determine whether the highlighter is on or off.

These considerations drastically reduced the number of potential interfaces. I then selected a set of 9 interfaces that would produce the most distinct coverage of the dimensions. This resulted in the interfaces described briefly below. Appendix I describes these interfaces in terms of the interface dimensions outlined above.

1. <u>Press Button Before</u>: To highlight text, the student clicks a button on a toolbar that floats above the text, and the next text they select is automatically highlighted. (see Figure 18)
2. <u>Press Button After</u>: To highlight text, the student selects text, and then clicks the toolbar that floats to the left of the text.
3. <u>Press Contextual Button</u>: When the student makes a selection, the toolbar appears at the top of the screen. The user can then click the highlight button to highlight text.

4. <u>Toolbar Follows Mouse</u>: When the student makes a selection, the toolbar appears next to the mouse. The user can then click the highlight button to highlight text.(see Figure 19)

5. <u>Press Key Before</u>: To highlight text, the user first presses the "h" button, and the next text they select is automatically highlighted.

6. <u>Press Key After</u>: To highlight text, the student selects text, and then presses the "h" button.

7. <u>Context Menu</u>: To highlight text, the user first selects it, and then right clicks on the selection. Upon right clicking, a menu pops up, and the user clicks the highlight entry to highlight the selected text.

8. <u>Click Selection</u>: To highlight text, the user first selects it, and then left clicks on the selection.

9. <u>Pickup Highlighter</u>: This interface mimics a traditional highlighter. Students pickup the highlighter by clicking on the highlighting button in the toolbar, which floats to the left of the learning materials. After clicking the toolbar, the button stays depressed until it is clicked by the user again. When the button is depressed the highlighter is "picked up", and every selection automatically turns into a highlight. (see Figure 20)

Students were also given a delete button, which could be accessed in several ways. The delete button could appear when the students either context clicked, left clicked, or moused over a highlight they had created. This was varied during the user-tests so that each interface was associated with different delete behaviors for different users.

Figure 19: Toolbar Follows Mouse interface. In this image, the student has finished making a selection using, and the highlight button pops up near the selection. Clicking the button turns the selection into a highlight. Note that this is a late version of the interface, where the button is always placed above the text in order to avoid obstructing subsequent text.

Figure 20: Pickup highlighter interface. The highlighting button is depressed in this image, meaning the highlighter is "picked up", or turned on. Every selection the student makes will be turned into a highlight. In this image the student is in the middle of a selection. Once the button is released, the selection will become a highlight. To put the highlighter down, the student clicks the highlight button again.

All interfaces could undo the previous highlight by pressing ctrl-z.

Only one of these tools explores the permanent side of the action duration dimension. As it is a permanent button-based interaction that occurs before the selection, the only other dimension to vary was the location of the toolbar. If this turned out to be important, it could be manipulated for the permanent interface in the second iteration of the study.

Initial User Tests

During the user-tests, each participant used three different highlighting interfaces. These were assigned so that the participants got experience with as many levels of each dimension as possible, so that their feedback with regards to these dimensions would be based on actual use. Six participants were included in the initial phase, meaning every interface was used exactly two times. The full assignment of interfaces to participants can be seen in Appendix I. Data from these tests that influenced the second round will be described here, while a further behavioral analysis will be reported below.

As in our previous study, a variety of selection-to-read behavior was encountered. Four of the six participants selected text frequently while reading, with no intention of using the selection. This behavior meant that contextual tools would pop up for actions that had nothing to do with highlighting. Though this would seem to be a

bigger problem for the condition where the toolbar follows the mouse, only one of its users did a very slight amount of selection to read (S4), and did not complain about the toolbar impeding their reading. However, this does point to the need to be very careful about where to place the toolbar when it follows the mouse, which will be discussed further below. Interestingly, the Pickup Highlighter tool impeded one heavy select-to-reader (S6) quite a bit. This user voiced strong complaints, and turned the tool on and off before each highlighting behavior, turning it into the equivalent of the Press Button Before interface.

Four participants were observed to occasionally follow the text they were reading with their cursor. This first became apparent when a student was using the cursor to track his reading and moused over a highlight, causing the delete button to appear. This provides an indication that mousing over highlights should not cause menus to popup, as it can often occur for spurious reasons. After observing these behaviors, the deletion dimension was reduced to either left or right clicking values.

With regards to ratings of the interface dimensions, four of the six participants preferred the buttons to be contextual. Only one person preferred the buttons to be always available, as it served as a reminder to highlight. Overall preference for contextual was 5.2 on the 7 point Likert scale. There was a strong preference for the action to occur after the selection is made. All subjects preferred this, and the overall Likert score was 6.8 on the 7 point scale in favor of after. In addition, several students (S2,S4,S5) using a tool that called for a selection after an action frequently tried to perform the action before the selection multiple times, doing the action after the selection seems intuitive.

Participants overall were neutral regarding whether the tool should be visible or invisible. Those who liked the visibility liked it because it served as a reminder. They were also neutral with regards to whether the tool should be permanent (3.16 on the 7 point Likert scale). With regards to location, 5 of the 6 participants preferred that the toolbar be located at the side of the text. This preference may be because a top location never impedes the text they are reading here, whereas a toolbar that floats at

the top will always impede text if the user needs to scroll, which occurred for at least one participant (S5).

Several other issues were observed with the interfaces. When participants did not like their initial selection, they would often click within that selection to create a new selection. The Click Selection tool then created a highlight, which was not the intention of the users (S5). Occasionally, there can be a delay between selecting text and highlighting, during which the user moves the mouse.

Interfaces for Second Phase

The second phase interfaces were intended to address questions raised during the first phase. The data described above suggests eliminating any tool whose action occurs before the selection, as participants strongly prefer this to choosing to highlight then selecting. However, they also lean towards permanent actions, which occur prior to the selection. As mentioned above, this preference may be due to the fact that most did not have an opportunity to use the Pickup Highlighter, and only one of those who did enjoyed the experience. Therefore the <u>Pickup Highlighter</u> was included in the second phase to gather more data regarding its use.

The <u>Toolbar Follows Mouse</u> tool was also included. It requires action after the selection and is a contextual tool, which is slightly preferred. It was also included with the hope that it would be used by more people who select-to-read, in order to determine whether and how it impeded reading by blocking text. Several positioning modifications were made during the testing in response to user observations, and will be described below.

The final interface included was <u>Press Button After</u>, as it also required action after the selection, and the toolbar was located on the left side. These tools both combined features from the Click Selection tool and Press Key After tools, as these tools were popular with the participants who used them. This collapsed the visible/invisible dimension. Students could either press "h" after selecting text, or double-click on the selection in order to create a highlight. Single clicking was

replaced by double-clicking in order to avoid the error described above, where an attempt to reselect text initiated the creation of a highlight.

Results

Five students participated in the second phase of user-tests, which followed the same procedure outlined above. Each student used all three tools, and order was counterbalanced. This means a version of the final tools was used by a total of 7 participants across both phases. In this section, I will first summarize results regarding the dimensions that describe the design space. I will then talk about specific implications for the design of highlighting interfaces. I will conclude by describing more general issues of both highlighting behavior and participants' descriptions of their goals and behaviors, which was the second thrust of this study.

	Responses to Interface Dimensions
1	*Timing*: Participants strongly preferred to press the highlight button *after* completing the selection.
2	*Duration*: There was a preference to "pick up" the highlighter. Potential issues include interruption of selection-to-read behavior (selections become highlights).
3	*Visibility*: Opinions were extreme, with some students liking visible toolbars as reminders of the interaction. Adding behaviors such as clicking selections to highlight appears to satisfy both sides, who can choose their preferred method.
4	*Contextual*: Again, ratings are extreme, though there is no average preference. A contextual toolbar that appears near the mouse when a selection is made is popular. Users report liking the reduced mouse movements.
5	*Location*: Probably dependent on the content. As selections, and with them the toolbar, often stay on the screen for extended periods of time. Participants are particularly frustrated when a toolbar obstructs text they may read
6	*Highlight Interaction*: Mousing over highlights should not raise menus, as students will often move the mouse in the course of reading. Clicking on highlights to access menus is a satisfactory solution.

Table 2: Responses to interface dimensions in the highlighting design space.

Interface Dimensions

Action Timing: Participants preferred to initiate the action after making the selection (5.8 of 7 Likert scale). As one user said "it felt backwards to turn on the highlight before; it made more sense to me to highlight and click" (S1). Another important point is that participants were not satisfied with their initial selection. "It doesn't make sense to do it before. What if you don't like the text you select yet?" (S3) As mentioned above, this desire to reselect also caused problems for the initial Click Selection tool.

Action Duration: Interestingly, students tended to prefer that highlighting last until the highlighter is "put down." (5.3). This may be due to the limited nature of the interface, as noted by participant 11: "I prefer for the highlight tool to stay active since it was the only tool offered in the module. The tool can be active since it is the only tool that we will be using repeatedly." Given more options, the cursor may

	Main Guidelines for Highlighting Interfaces
1	Toolbars should not obscure text students may read. As selections may stay active for extended periods of time, contextual toolbars should be placed *above* the selection.
2	Highlights should not have mouse-over behavior, as it will be invoked for spurious reasons such as when users follow their reading with the cursor.
3	Clicking on a selection should not be the source of an action like highlighting, because it often occurs when students are attempting to improve their original selection.
4	Combining visible (toolbars) and invisible (keystroke) actions provides multiple pathways.

Table 3: Design guidelines for Highlighting interfaces

become hopelessly overloaded. In addition, some students strongly prefer the action to occur for only one selection, "because you select text sometimes when you don't decide to highlight them" (S2). Picking up the tool requires precise selection.

Visibility: While overall ratings are neutral on the visibility scale (4.3), ratings tend to be on the extreme side. For some users a visible toolbar serves as a reminder to highlight material. "I like seeing it as a reminder" (S1). Those who like the invisible interfaces often liked them for ease and reduced mouse movements. Students especially liked Click Selection because "it is the most natural and easiest for me" (S5) and it required "less movement of mouse left and right" (S3). As demonstrated in the second phase of this study, *an interface can provide users with both visible and invisible paths to highlighting.*

Contextual vs. Permanent: Overall ratings were also neutral with regards to whether a toolbar should be permanent or contextual (4.2), though once again opinions were extreme. As above, some participants "liked having the toolbar visible as a reminder of the tool being there" (S1). Others felt "why do I need it when I am not selecting" (S2). It should also be recalled that permanent actions can impede other behaviors such as selection-to-read. One design consideration is whether the placement of a permanent tool is likely to obscure text when users scroll, as was the case with a toolbar at the top of the content in this experiment. The next section will describe how these issues were dealt with for the Toolbar Follows Mouse interface.

<u>Location</u>: Eight of the 11 users preferred that a toolbar be located at the side of the text rather than the top. This is likely a factor of the webpages used, as they had blank space on the left so the toolbar never obscured text on the left side. However, a toolbar at the top of the window will always obscure text unless it is part of the overall browser toolbar. Eight of the 11 users preferred that a toolbar be located next to the mouse than on either the side or the top if the toolbar was contextual, only appearing when a selection was made.

Positioning of Contextual Toolbar

The positioning of the toolbar in the Toolbar Follows Mouse was subtly manipulated several times in response to observations during this study. The implementation of events in the browser used meant that there was often a delay between the termination of the selection event and calling the mouseup. This meant the toolbar could be placed at a distance from the selection. Users often moved the mouse a distance before deciding to highlight, which meant the toolbar was not located near either the selection or the mouse. This occasionally made it difficult for the participant to find the toolbar when they decided to highlight. Therefore instead of being positioned next to the mouse after a selection was made, the toolbar was placed next to, but not covering, the selected text at the shortest distance possible from the mouse.

However, now the toolbar would occasionally block the next line of text, which the students could not then read. This was caused most frequently when participants were selecting-to-read. It also occurred when participants were selecting a piece of text they were considering highlighting, but first read the following line to clarify their thinking. Therefore the final solution was to <u>place the toolbar directly *above* the selected text, inline with the mouse</u>. While the toolbar would still obscure text, it would reduce it to the less frequent situations in which readers are looking backwards in the text.

Design Recommendations

The Toolbar Follows Mouse was the clear winner if students are to be given only one visible method of taking notes. Out of the final candidates, it was the winner with regards to user satisfaction. Across the phases it had a 6 rating on the 7 point Likert scale, and no one disliked the interface. The other two tools had a 5.1 rating (4 was neutral), and ratings were more extreme for these tools.

Though the interface obscures text, it does not appear to interrupt selection-to-read behavior, especially when its positioning was refined as described above. In addition, participants preferred the toolbar be located next to the mouse rather than at the side of the page, as this reduced mouse movements. The tool should also include other invisible ways to quickly create a highlight, such as double-clicking the selection or pressing a key.

While it appears that all users will be satisfied with this tool, other participants may prefer a different interface. However, as the other interfaces were strongly disliked by some users, none should be the only option available. In particular, many participants enjoyed the Pickup Highlighter interface, as it allowed them to easily highlight many different pieces of text. However, as mentioned earlier, this was frustrating when it got in the way of selection-to-read behavior. Some readers may prefer the permanent presence of the toolbar at the side of the text as a reminder of the functionality. One could imagine an interface that allows readers access to all of these interfaces, with the ability to attach and detach the toolbar from the mouse, and select a permanent or temporary highlighter. Designing the process of customization could be quite tricky and involve quite a few design decisions. There are multiple dimensions on which the design would have to be customized.

With regards to deletion, as described above, hovering was eliminated as an option because it cause the delete toolbar to appear when readers were following the text with their mouse. Participants slightly preferred left-clicking on the highlight to right-clicking to access the delete toolbar. This may be because right-clicking already has a function in this context (opening the context menu), whereas left-clicking does not.

Left-clicking on the highlight appears to be a good method of accessing highlight menus, which for more advanced annotation devices could include comments.

Behaviors and Reports

Participants took notes at different times in the reading process. The most common was to highlight text after they had finished reading the sentence, occasionally while they were reading a subsequent sentence. Other participants started the highlighting process in the middle of reading a sentence. Several students were observed to highlight text before they actually read the text (S5, S6, S7). This occurred for text whose importance was signaled by either bolded words, or a header that identified the text as a definition. Students were also observed to read text, highlight it, and then reread it, as if they had identified the text as important but were trying to understand it in greater depth.

Students occasionally verbalized what led them to record a note. This was most often a vague statement that "this is important information" (S1, S2, S3, S11). But others stated they were recording a good example (S1, S6), something that confused them (S2, S9), a good summary (S1, S6), or something that was difficult (S10), or something that helped them understand material (S6).

Goals and Motivations

The semi-structured interviews were used to obtain information on why students highlight materials, as well as why they feedback on other behaviors. As in the previous study, selection-to-read, which was observed in 7 of the 11 participants, seems to be a function of paying attention. It "keeps me awake and actually doing something" (S3). Interestingly, one participant reported something that sounds quite a bit like Marshall's "narrowing of focus" described in the previous study (Marshall & Bly 2005). For this user selection-to-read helps solve the problem of "when I look at the whole page, it's hard." (S9). Though 8 of 11 participants were observed to at least occasionally follow text with their mouse while reading without selecting, a failure of the interview is that it did not touch on this behavior with many participants, so it is a difficult behavior to interpret. One of the participants who did

give feedback regarding mouse-following believed it was an artifact of the requirement to think out loud (S5). Another reported that they actually follow text in the book with a pencil when they are reading (S10).

Participants gave us a variety of reasons for highlighting text, some of which are familiar to the responses observed for note-taking in general. All participants say they highlight material in order to facilitate review, both with regards to speeding up the review process and focusing review on the critical components. Three participants stated that the act of highlighting text facilitated learning.

The interview also dealt with the different types of notes taken. As above, students report highlighting definitions they believe they need to know as well as important examples that help them understand the material. They also report highlighting structural elements such as headers and titles. They distinguish between highlighting individual terms and key phrases or sentences, though it is not clear from the interviews the different functions these play. There appears to be a distinction between things they believe they need to know for testing, such as definitions, and material that helps them understand the concepts, such as examples or supporting ideas.

Summary

One of the more interesting results of this study was the finding that for some interface dimensions, while average preference was neutral, individual preferences were extreme. In the case of the visible/invisible contrast, both sides could be satisfied by integrating invisible features such as clicking selections to highlight them with visible toolbars. There is no obvious solution to others, such as the contextual/permanent distinction. However, investigating students' actual behavior and specific tool preferences indicated that even users who preferred permanent interfaces liked the "Toolbar Follows Mouse" interface. Where the interaction between multiple dimensions can be difficult to tease out, as in this case, it is important to test possible design solutions, so putting the "Toolbar Follows Mouse"

tool in front of more users was useful. Another solution suggested above is to give users control over controversial dimensions where possible.

In this work I refined the design study process described in this thesis to be more systematic in how the design space was explored, how initial interfaces were developed, and how interfaces were assigned to users. The addition of surveys that asked students to state preferences with regards to interface dimensions was also useful. While this is still not meant to be an experimental test of the design space, the triangulation between stated preference and behavior is useful.

Chapter 11: Highlighting vs. Copy-Paste

While the design study described above was intended to provide insight into the design space of highlighting applications, it was motivated by the need to develop a highlighting interface that could be experimentally compared with the copy-paste note-taking interfaces from the experiments described earlier in this thesis. Here I report a study comparing highlighting and copy-paste note-taking.

This study investigates three high-level questions. First, are copy-paste note-taking and highlighting similar with regards to behavior and learning outcomes? Second, does the presence of a notepad play a role in learning outcomes and how students record notes? Finally, the experiment investigates students' perceptions of the similarities and differences between note-taking and highlighting.

Interfaces Included

The following interfaces were evaluated in this study:

Copy-Paste: This tool allows students to record notes into their notepad by copy-pasting or dragging and dropping selections from the text. The notepad takes up the bottom third of the browser window. Students cannot type in the notepad, but they are offered basic markup functionality (bold, italic, underline), outlining (bulleted lists, indenting, outdenting), and editing (deletion of text, dashes, parentheses, and spaces).

Highlighting: This tool allows students to highlight textual material. When students make a selection, a small toolbar pops up above the selection. The student can highlight the selection by either pressing the button, pressing the "h" key, or double clicking on the selection. The previous highlight can be undone by pressing ctrl-z, and any highlight can be deleted by first clicking on the highlight, and then clicking on the delete button that pops up.
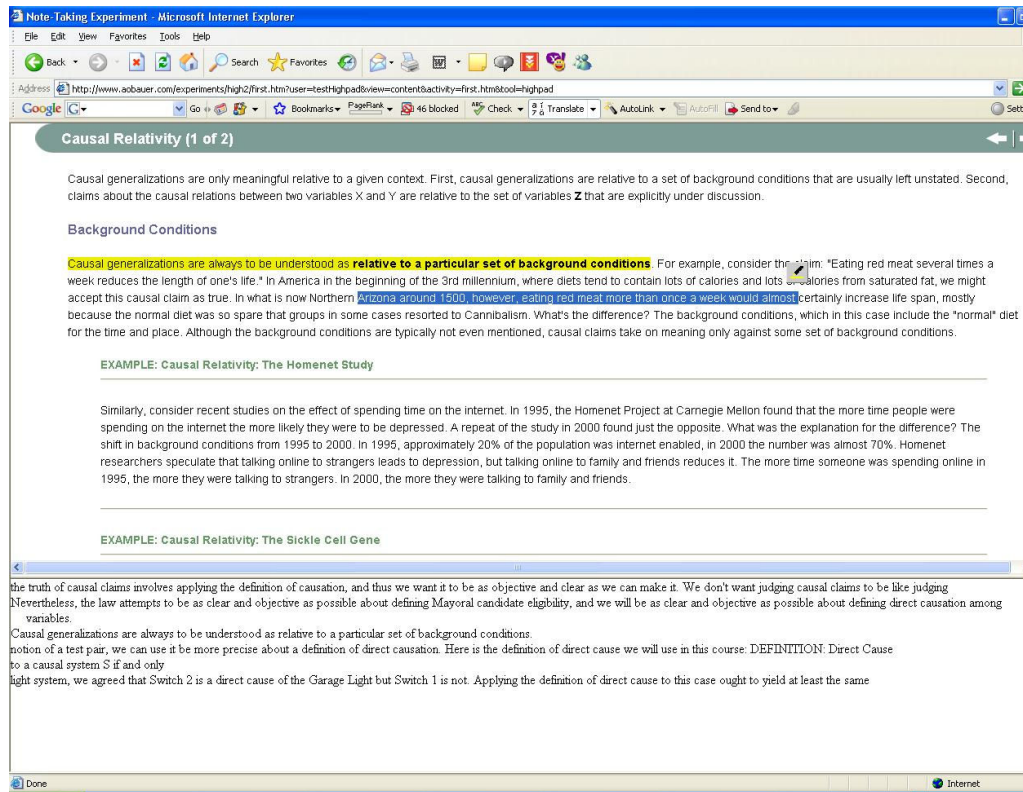
Figure 21: Highpad interface. After selecting text, a button appears above the selection that allows students to highlight the text. Once text is highlighted, it is placed in the notepad viewed at the bottom of the screen in the order it appears on the page. So if the current selection was to be highlighted, it would be placed in the notepad after "Causal generalization as are always…" Students cannot edit or reorganize materials in this notepad.

Highlighting with Notepad (Highpad): This tool adds a notepad to the highlighting interface described above. The notepad is located in the bottom third of the screen. Whenever content is highlighted, it is automatically added as a new paragraph in the notepad. The content in this notepad cannot be marked up or reorganized. Organization follows that of the learning material, so if both the first and last sentence on a content page were highlighted, the first one would appear before the last one in the notepad. The contents of the notepad cannot be modified except through addition and deletion of highlights. Students cannot edit, markup, or reorganize the material. (see Figure 21)

No Notes: In this control condition, students are not allowed to take notes of any kind; they were simply asked to read through the learning materials.

There is a question regarding how the Highlighting condition should be presented with their notes. In the Copy-Paste and Highpad conditions, students were actively constructing the notes they would review, whereas in the Highlighting condition students were not creating any visible document. When highlighting in pencil and paper, readers can only review the highlighted material in the context of the whole document. However, this would require all conditions to be given the entire content to review, which is not traditional to note-taking studies and would require more time to review. Presenting Highlighters with the same style of notes as Highpad allow us to determine whether extracting the highlighted material can be useful to students.

In all conditions, clicking on a sentence selects the entire sentence. The inclusion of this feature was popular in the previous study, actually increasing user-satisfaction over the unrestricted tool even though the interface in that study restricted selections to single sentences. This study does not include the single-sentence restriction. As the restriction was not shown to influence learning outcomes, there is not a strong rationale supporting its inclusion.

Hypotheses

1: <u>Copy-Pasting, Highlighting, and Highpad will perform better on learning outcomes than the No Notes Condition</u>. Note-taking has been shown across a variety of studies to increase learning. Though this result was not confirmed in the previous study, the materials were altered slightly as described below in order to re-evaluate this hypothesis.

2: <u>Tools with a notepad will perform better on learning outcomes than the highlighting tool</u>. As described above, a notepad may allow students to coordinate what they are learning with what they have already learned. This should, in terms of models of reading, strengthen their textbase, which will in turn increase performance on learning outcomes.

3: <u>The ability to manipulate notes will allow the Copy-Paste condition to perform better than the Highpad Condition</u>. Previous research into note-taking has found that students who reorganize their notes perform better on learning outcomes than students with static organization (Shimmerlik & Nolan, 1976). In that experiment, students were required to reorganize, while in this study the copy-pasters had the ability to reorganize and edit their notes, but were not forced to do so. Still, as only the copy-pasters were able to manipulate their notes and editing behavior had been observed anecdotally in previous experiments, they were expected to perform better on learning outcomes.

4: <u>Copy-Pasting and Highlighting will show similar selection behavior and learning outcomes linked to those behaviors</u>. As highlighting and copy-pasting share similar interaction techniques, we expect them to be linked with learning outcomes in similar ways. The important example is that of selection size. The previous studies linked increased selection size and reduced learning for copy-pasting, and I expect this result to transfer to highlighting.

The questionnaire portion of this study addresses students' perceptions of the hypotheses outlined above. In addition, it evaluates student's perception of any differences between note-taking and highlighting and reasons behind selection-to-read behavior.

Subjects and Materials

This study followed the same between-subjects design of previous experiments. A total of 54 subjects from several local universities were recruited by means of a posting to a subject-recruitment website. Three subjects were not able to complete the materials; their data was not included in the analyses described here.

Each quiz contained 23 items, which targeted the 10 ideas around which the instructor based the module. 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.

The previous study did not find strong benefits for note-taking over simply reading the materials. The specific content could be behind the lack of effect, as note-taking effects are not seen across all studies (Kobayashi 2005, 2006). Unfortunately, attempts to obtain different materials that had shown note-taking effects were unsuccessful, as recent studies that found effects for different note-taking interfaces or techniques did not include no-note control conditions (e.g. Igo et. al. 2005), so it is unclear whether those studies actually found overall note-taking benefits. Instead I attempted to improve the tests and learning materials from the previous study. An item analysis was performed on the data from the previous experiment evaluating the difficulty and discrimination of each test item. New test items were substituted for old items with low discrimination. We used the think-aloud from the design study to identify and replace confusing elements from the learning materials. This consisted mostly of removing text that referred to quizzes that had previously been removed.

Behavioral data was collected in the same manner as in the previous study. As all notes were verbatim, the same Excel VBA macro could be used to associate highlights and copy-pasted notes with their respective sentences in the test. After being split into sentences, the notes were coded with regards to whether or not they were alone in the action (e.g. whether a sentence was highlighted or pasted alone, or whether they were pasted along with other selections) and wordiness, or the percentage of possible words were actually recorded.

Survey

The survey students were given at the end of the experiment was aimed at evaluating students' experience using the interfaces and answering several questions regarding student goals and behaviors.

Students were asked several questions regarding user experience. First, they were asked whether they would use the interface in an actual online class. They were then asked to report the three favorite and least favorite aspects of the interface they used. Further questions asked participants to answer 3 questions on a 7 point Likert scale, regarding the pleasantness of the tool, whether it allowed them to accomplish their goals, and whether they thought the tool helped them learn. Students in pasting and Highpad conditions were asked whether they referred back to their notes, and students in the Highlighting condition were asked whether they found their notes useful for review.

Participants were also asked several questions regarding their conceptions of highlighting (defined as highlighting, underlining, or circling) and note-taking (handwritten or typed on a piece of paper). They were first asked to rank their goals by importance both when highlighting and when taking textual notes. They were given 5 options, taken from the literature (e.g. Van Meter et. al., 1994) and the design study interviews: paying attention, the process helps learn, to review quickly, to review important materials, and to share with others. Participants were given space to fill in and rank additional goals.

Participants were also asked to rank the type of content they were most likely to record in notes or highlight. Again, categories were taken from the literature and the design study interviews. The categories were: definitions, key words/terms, key ideas or phrases, good examples, text that helps understand, text that is confusing, and good summaries. Students were again given space to fill in and rank additional types of content.

Results

Pretest was found to be a significant factor in both time on task and learning outcomes, so was included as a covariate in a full factorial with interface condition in the analyses of these outcomes. Native language was also found to be a significant factor in time-on-task, so it was included in this analysis, though it was not significant for any other outcome, nor did it interact with any other factor.

ANOVAs were performed on time on task, with the factors described above. There was a significant overall effect of condition with regards to time on task $F(3,42)=2.84$, $p<.05$, language (binary-whether English was the participants native language)

Time

Figure 22: There was a significant effect for time on task. The Highpad condition completed the module significantly faster than other conditions.

native language) $F(1,42)=5.1$, $p<.01$, and pretest $F(1,42)=9.06$, $p<.01$ (see Figure 21). Contrasts indicate the Highpad treatment was significantly faster than all other conditions, which were not significantly different from each other. Students using the Paste tool spent on average 10 minutes, or 20% of their total time, interacting with the notepad. If this time is subtracted for each copy-paster, the treatment no longer performs significantly slower than the Highpad tool. Though it then trends below the other two treatments it does not perform significantly faster than them.

Native English speakers completed the module 17% faster than non-native speakers, and students who scored high on the pretest completed the module 24% faster than students with low scores on the pretest. There was also a marginally significant treatment by pretest interaction $F(3,42)=2.46$, $p=.07$. While both the Highlighting and the Highpad conditions completed the module significantly ($p<.005$) faster when they came in with more knowledge, this was not the other case for the other treatments.

Learning

ANOVAs were conducted on each test, with treatment (Paste, Highlight, Highpad, and No-Notes) and Pre-Test Mean Split (Hi, Lo) included in the model in a full factorial. A marginally significant effect was found for the review free response test $F(1,43)=2.18$. $p=.1$. While the Pasting condition performed significantly better than No-notes ($p<.05$) and the Highpad condition performed marginally ($p<.1$) better than no notes, Highlighting did not perform better. No other significant or marginally significant effects were found on individual tests.
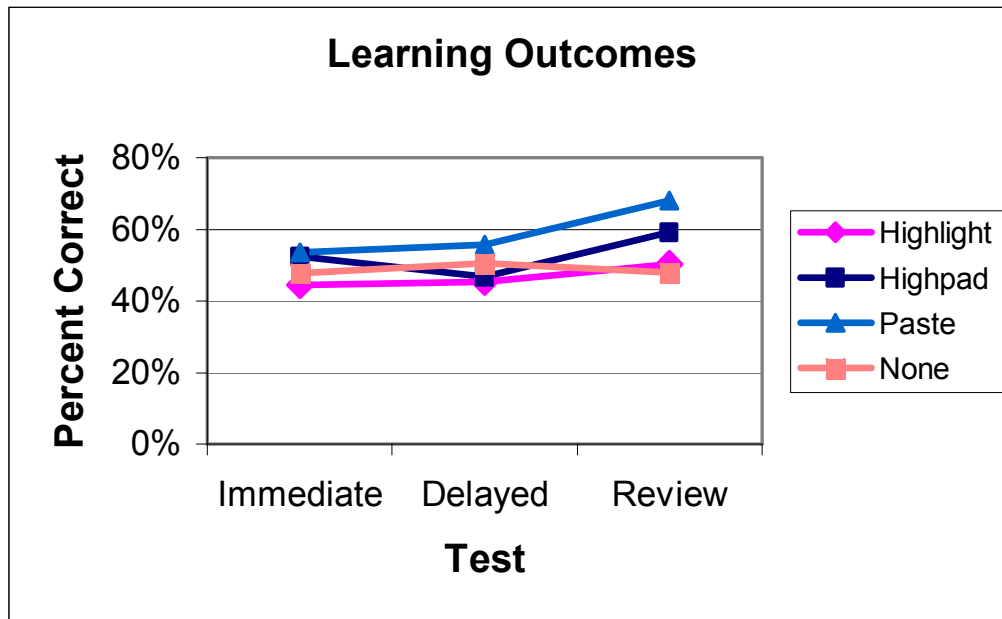
Figure 23: Learning Outcomes. Copy-Pasting performed significantly better on free response items on the review test, and Highpad performed marginally better. Both received significant benefits from review, while Higlighting and No-Notes did not.

Repeated measures analyses were also performed across all tests, searching for condition by test time interactions, with pretest in the model as in a full factorial. This would measure forgetting and remembering between tests. Test time was significant $F_{(2,42)}=5.78$, $p<.01$, and there was a marginal condition by test time interaction $F_{(6,84)}=1.84$, $p=.1$. Exploring this interaction found that there was not a significant time or time by condition interaction for the immediate and delayed tests. However both time $F_{(1,43)}=11.4$, $p=.001$, and time by condition $F_{(3,43)}=3.2$, $p<.05$, were significant between the second and third test. Both the Highpad and Copy-paste benefited from review, while highlighting and no-notes did not (see Figure 23).

As time differed, we also looked at efficiency scores on each test, using the metric developed by Paas and colleagues (Paas et. al., 2003). An efficiency score was calculated by subtracting the z-score of test-performance from the z-score of time and dividing by the square root of two. In an ANOVA with condition and Pre-Test in a full factorial, condition was marginally significant for the immediate test $F_{(3,43)}=2.2$, $p<.1$, and significant for efficiency on the final test $F_{(3,43)}=3.49$, $p<.05$. On the first test, the Highpad condition was significantly more efficient than the other conditions, which were not significantly different. On the final test, the Highpad condition was more efficient from all but the Paste condition, and no other conditions were significantly different from each other.

Behavior

ANOVAs were also conducted with measure of note-taking behavior with condition and pre-test in the model. Pre-test did not interact with condition on any measure. Pre-test was only a significant factor with regards to total number of ideas recorded $F_{(1,35)}=4.03$, $p<.05$. Students who knew more coming in made fewer highlighting or copy-paste actions.

There were significant condition effects for the total number of sentences recorded $F_{(2,35)}=3.07$, $p=.05$ and the percent of sentences students recorded by themselves

F(2,35)=3.51, p<.05. Students recorded more sentences using both the Highlighting interface and the Highpad interface than they did when copy-pasting. (see Figure 24) Students recorded a higher percentage of sentences individually using the Highpad tool than they did using the other tools. (see Figure 25) Students did not record a different amount of

Number of Ideas

Figure 24: Total Ideas and Key Ideas. The two highlighting conditions record significantly more total ideas than the past condition. The conditions do not differ with regards to the number of key ideas recorded.

key ideas F(2,35)=.77, p>.4. Wordiness did not differ either F(2,35)=1.33, p>.2 with regards to key idea.

We also collected selection to read data for students in the No Notes condition. All but three of the 12 subjects either selected blocks of texts or clicked to select on at least half of the pages. Three of 12 the subjects displayed selection-to-read behaviors more than twice a page.

The previous studies in this thesis found connections between notes and learning. In this experiment, associations between presence in notes and performance on test were not possible, as there were few instances in which a key idea was not recorded. This was not the case with regards to whether the idea was recorded in a selection in which it was the only idea present. ANCOVA was performed

Ideas Recorded Alone
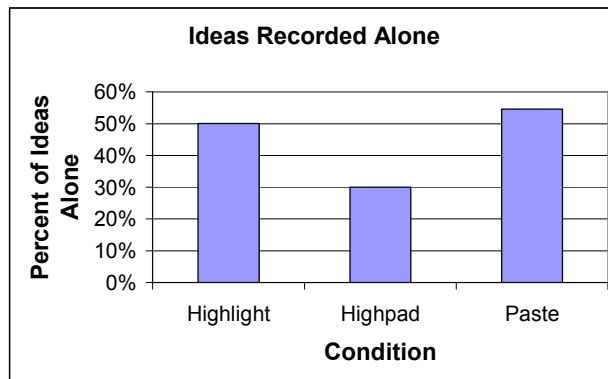
Figure 25: Percentage of Ideas Recorded Alone. The Highpad tool records significantly fewer ideas by themselves than do the other two conditions, which do not differ from each other significantly.

with item correctness as the dependent measure, pre-test and "ever-alone" as covariates, treatment as a between-subjects variable and test-time (immediate vs. delay vs. review) as a within-subjects variable. There was no significant effect for "ever-alone" $F(1, 2397)=.01$, $p>.9$, and it did not interact with treatment $F(2, 2397)=1$, $p>.3$.

Stated Preference

Participants' reactions to the tools were recorded in the questionnaire. While three-fourths of students in the Highlighting and Pasting condition said they would use the tool in an actual class, less than half of students in the Highpad condition said the same. There was a significant difference with regards to pleasantness of the interfaces $F(2,36)=4.69$, $p=.01$. The Highlighting condition was significantly preferred over the Pasting and Highpad tools, which did not differ according to preference. There was also a significant difference with regards to whether students thought the interfaces helped them learn $F(2,36)=4.69$, $p=.01$. Students thought the Highlighting and Pasting tool helped them learn more than the Highpad tool, which on average they did not think helped them learn. There was no significant difference with regards to whether the interface allowed the users to accomplish their note-taking or highlighting goals $F(2,36)=1.87$, $p>.1$.

As in previous studies, most students (two-thirds) in the Pasting condition wanted the ability to type. With regards to the Highpad condition, a third of participants wanted the ability to type, while half wanted the ability to reorganize their notes by making outlines and moving around text. Six of the 27 students in the highlighting or Highpad conditions wanted the ability to choose additional colors, and 3 more wanted to be able to underline in addition to highlighting.

Ten of the 12 students in the Highpad condition found the notepad to be useful while they were taking notes. Five stated that they referred back to notes they had taken on previous pages, while eight of the 12 students in the Copy-Paste condition reported doing so. Two-thirds of the students in the Highlighting condition found

Figure 26: Students do have significantly similar goals when taking notes and when highlighting. Students do not believe they learn from the actual process of highlighting, using the technique mostly to help them review.

the notes they were provided with useful for review between the delayed and the review tests.

Note-taking vs. Highlighting

Students were asked to compare note-taking and highlighting in the questionnaire. Ratings were standardized for each student. Pairwise correlations were performed on all Note-taking/Highlighting pairs for which both were ranked. There were situations in which a student ranked an item for one condition and not the other, and these were also eliminated.

The first question asked students to rank the importance of different motivations for taking notes. A total of 35 of the 255 (14%) pairs were taken out of the data for having at least one blank value. Nine of these were blank for both values. The possible values were: Attention, Processing Benefits, Reviewing Quickly, and Reviewing Important Ideas. There was a weak positive correlation $r(218) = .16$, $p =$
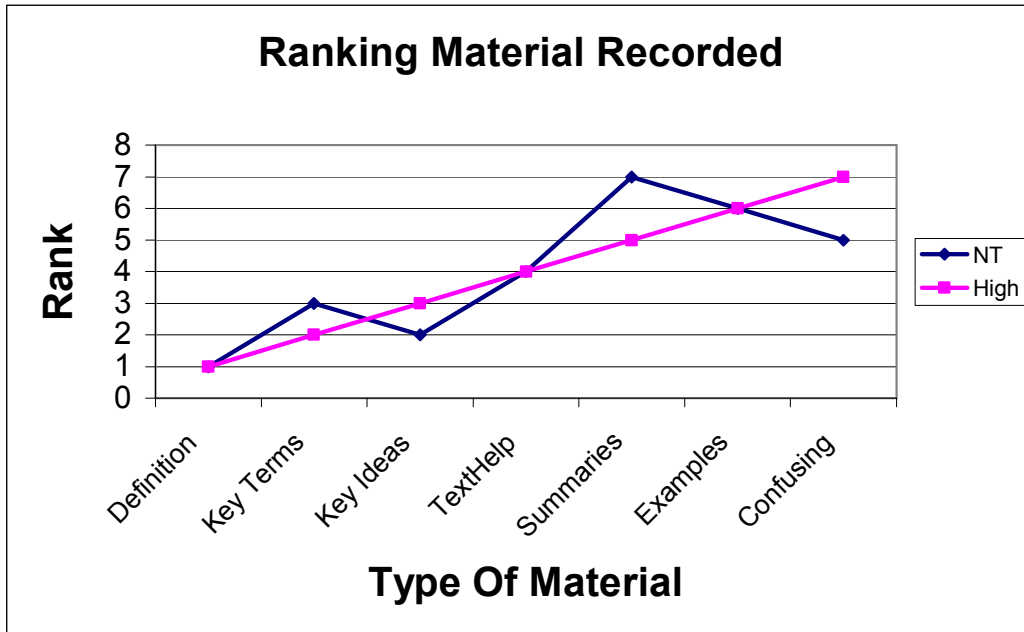
Figure 27: There was a significant correlation regarding what material students focused on when highlighting and when note-taking. The average rankings across students were also correlated. Students seem to focus in mostly on definitions, key terms and key ideas.

.01 between how students rated their motivations for note-taking and for highlighting.

Figure 26 shows how students on average ranked their different goals. The rankings of average standardized scores are not correlated r(3)=.13, p>.8, though the data is of course limited. Students' top reason for using note-taking is for its process benefits, believing that the act of recording notes helps them retain information. This is least important for highlighting. Highlighting is used more to allow students to quickly review information.

Students were also asked to rank the types of material they were likely to highlight or record in notes. A total of 49 of 357 (14%) pairs were taken out of the data for having blank values. Nine of those these were blank for both values. There was a moderate correlation between Note-taking and highlighting for the material students reported taking notes r(199)=.39, p<.0001.

As seen in Figure 27 average ranking is similar across the two. Though again, data is limited, the ranking of average standardized scores is highly correlated r(5)=.9, p<.001. In both note-taking and highlighting, students concentrate mostly on recording definitions and key terms and ideas.

Discussion

Learning Hypotheses

We did not find strong evidence for the first hypothesis, that both note-taking and highlighting would perform better on learning outcomes than not taking notes. As there were no significant results on the first two tests, neither note-taking nor highlighting provided an **encoding** benefit.

There is limited support for the second hypothesis, that the presence of a notepad will benefit learning. Both the Highpad and the Copy-Paste treatments performed better on the review test for free response items than the No-Notes treatment. Therefore there does appear to be an **external storage** effect for conditions in which students have access to their notepad while they are highlighting or copy-pasting.

There is no evidence for the third hypothesis, that manipulating notes in the Copy-Paste treatment will result in superior learning relative to Highpad treatment, where the notepad cannot be edited.

Note-Taking and Highlighting

In previous studies I found that copy-pasting was more efficient than typing, producing similar learning results in less time. In this study, I found the Highpad treatment produced more efficient learning than the other tools, performing equivalently or better to all other conditions in the least amount of time. There is some evidence that highlighting interaction in general allows high knowledge students to skim the material they are reading. Students who performed well on the

pretest finished the module more quickly than low knowledge students when using the two highlighting tools, though this was not the case in other treatments.

The Highpad treatment completed the module significantly faster than the other conditions. Learning was more efficient for the Highpad treatment on the immediate test. It was also more efficient than all but the Paste treatment on the final test. One of the main differences between the Highpad and Copy-Paste condition was the ability to interact with their notes. Students using the Paste tool spent approximately twenty percent of their time editing and manipulating their notes, though they received no benefit with regards to learning or review. This provides additional evidence against hypothesis 3, that manipulation of notes improves learning outcomes.

Students recorded more sentences in both conditions in which they were given the ability to highlight. However, they covered the same number of key ideas. Students using the Highpad tool were much more likely to highlight multiple sentences in one selection than either Highlighters or Copy-Pasters. However, none of these behaviors was linked with learning outcomes.

Students seem to view highlighting and note-taking in different ways. Within students, there is only a weak correlation between the motivations underlying the two activities. The major difference regards process benefits, or the feeling that the act of recording a note or highlighting text increases retention. This is the most important component of note-taking and least important component of highlighting. Students using highlighting are more focused on hastening the act of review.

Students focus on similar types of content when highlighting and note-taking, and on average the rankings are strikingly similar. Recording definitions is most important to both styles, and while highlighting places more importance on selecting key terms, note-taking focuses more on the larger scale of key ideas.

Implications for Design

The highlighting tool designed for this study was the preferred interface in this experiment, rating a point higher in the 7-point Likert scale of pleasantness of use. Not a single user described it as unpleasant. Most students found the addition of a notepad in the Highpad condition useful, and it allowed them to create notes that improved performance on the review test in the same way observed in the Paste condition. It may be that the presence of a notepad makes the process more like copy-pasting, only without the time sink of organizing and editing the notepad. In addition, students using this tool were more efficient than any of the other conditions, achieving similar or superior learning in less time.

However, students liked the tool less, and most stated they would not use it if it were provided in an online class. Being able to see their notes caused the students to want to reorganize and edit them. As seen in the Paste condition, the ability to do so significantly increases time on task. Allowing Highpad users to edit their notes may eliminate the efficiency benefits observed in this experiment.

While students have different reasons for taking notes, they appear to focus on the same type of material. Though students in general do not believe the process of highlighting helps them learn, adding the notepad may allow them to integrate the two. According to both their desires and the desires of students in the Paste condition, however, adding organization and editing functionality will be important.

Alternative designs could be explored. The notes created through highlighting could be available in a popup, which students could access at any time. Knowing it is there may cause students to highlight in the effective ways seen by Highpad users, while its lowered salience could reduce the amount of time students spend interacting with their notes.

Summary

This study evaluated whether a notepad plays a part in facilitating learning by allowing students to coordinate what they are learning with what they have learned. We did not find any evidence for such coordination, though having a visible notepad allowed students to produce notes from which they could review. In fact, students spent a significant amount of time editing their notepad in the Paste condition, but as this did not result in performance gains, this could be viewed as wasted time.

There are interesting similarities and differences with regards to how students view note-taking and highlighting. Behaviorally, even copy-paste based note-taking and highlighting are different. Highlighting appears to result in uninhibited note-taking, as both conditions in which students could highlight recorded significantly more total ideas. Interestingly, adding the Highpad did not make highlighting behaviorally similar to pasting, though both got identical review benefits, which were not observed in highlighting only.

Students do not have the same motivations for note-taking and highlighting, though they focus on the same materials. Students do not place much importance on process benefits of highlighting, whereas that is the most important feature of note-taking for them. The importance of processing to note-taking may be why they spend so much time organizing their notes, though they do not appear to gain any process benefits from doing so.

Chapter 12: Conclusions and Limitations[CMU1]

The research described in this thesis was centrally motivated by the desire to understand the influence technology has on both students' note-taking behaviors and the learning outcomes achieved through note-taking. The results reported here reinforce the importance of this type of research. **Not only does the design of technology built to support digital note-taking affect how students take notes, but different interfaces can produce different learning outcomes.** While there are a variety of note-taking applications in existence, which have produced demonstrably different note-taking behaviors, most lack empirical evaluations of learning outcomes. It could be that the interfaces being developed are actually detrimental with regards to learning. This research provides a first step in addressing this problem.

Note-taking has a rich history of research, which has produced a range of theories regarding the mechanisms underlying the relationship between note-taking and learning. This thesis presents a series of studies, intended to extend past literature in various ways. I will describe how the work I have reported here contributes to past work, and then outline some of the more specific contributions this work makes with regards to knowledge about the design of note-taking applications.

Note-Taking and Learning

Encoding and External Storage

Both the process of recording notes and having notes for review have been shown to promote learning. My studies have found only the slightest hint of processing effects, where the inclusion of copy-paste functionality appears to increase forgetting. Students using different interfaces never performed significantly differently on either an immediate or delayed post-test.

As is true in the literature at large, in the studies reported here **review is a much more robust effect of note-taking than processing**. All interfaces but the novel interventions from study 3 and the highlighting-only interface received review benefits. With regards to the interventions from study 3, review benefits were most likely eliminated because students recorded fewer key ideas. In the case of highlighting, students apparently were not able to produce notes from which they could successfully review. However traditional highlighting allows students to review a marked up document, whereas in the study reported here students could only review the contents of their highlights.

Focusing and Elaboration

The note-taking literature identifies focusing and elaboration as two key contributions note-taking makes to learning. Several studies in this thesis addressed questions of focus. The negative association between wordiness and learning for the copy-paste tool found across multiple studies could be a sign that students are not focusing on the key features of the learning materials, as copy-pasting does not even require students to read the material. However, interventions that effectively encouraged small selections without reducing satisfaction or use did not realize learning gains, so intervening to reduce selection size does not appear to increase focus.

There were two tests of elaboration. Students are often encouraged to elaborate by recording notes in their own words. These studies do not support this recommendation, as students gained no benefit from rewording their notes. However, students were not observed to bring outside knowledge into their notes. Models of reading comprehension suggest that connecting outside knowledge with readings results in a strengthened situation model, which in turn increases performance on measures of learning transfer. **It may be that students should not be asked to simply reword notes, but to add relevant outside information to their notes.** Simply rewording reduces the efficiency of note-taking without increasing retention. While it is not clear that students reword notes with the

expectation that this will help them retain information, students do report a belief that the verbatim note-taking produced by copy-pasting is detrimental to learning.

The second elaboration hypothesis evaluated here was whether a notepad allowed students to coordinate what they are currently reading with what they have previously read. I hypothesized that the availability of multiple sources would result in bridging inferences, strengthening the students' textbase, and increasing performance on learning outcomes. While students did report referring back to notes from previous pages, this behavior did not result in increased performance on learning outcomes relative to students who did not have access to notepads. The simple presence of a notepad only increased performance on review tests, where they were given access to the notes they had created a week earlier.

Learning Efficiency

One of the more interesting findings regarded **learning efficiency**. Though interfaces often did not differ with regards to learning outcomes, they did differ with regards to time on task. This meant that some interfaces resulted in students learning the same amount in less time. In particular, when students can only copy-paste, they finish the module more quickly than note-takers who can type, and perform equivalently on learning outcomes. In addition, highlighters with access to a notepad learn the same amount in less time than copy-pasters or students who can only highlight material. Students spend quite a bit of time editing their notes, with no observable learning benefits. This is an important result, because as described in the introduction, often when note-taking is found to benefit learning, it also increases time on task.

Note-Taking Behavior

My research suggests that the functionality students are provided with in an interface changes how they record notes. In particular students will use the copy-paste functionality to produce far more notes, that are of a more verbatim and wordy nature than if they type or handwrite notes.

The literature identifies wording as an important feature of note-taking. Many researchers believe that note-taking is superior when students record ideas in their own words, though the empirical evidence is equivocal. **The results here do not find a learning benefit of own wording.** While different interfaces produced different wording, they did not differ with regards to learning. More fine grained analyses attempting to connect wording of key ideas with performance on learning outcomes also failed to find a significant effect. It is important again to mention that students were not observed to use their notes to connect the learning material with outside information. It may be that simply rewording ideas is not sufficient to improve students' encoding of the learning material.

This research did find effects of wordiness. Wordiness is also noted in the literature as an important indicator of the quality of note-taking, though again the evidence is not conclusive. My research provides some evidence that the relationship between wordiness and learning may be a function of the cost of wordiness. When increased wordiness is not costly, as is the case with copy-pasting, it was actually associated with negative learning outcomes. Wordiness was either positive or neutral for more costly interfaces such as typing or handwriting.

Intervention

Traditional note-taking research has found that pretraining and instruction are ineffective in encouraging learning, partly because many students do not comply. Behavioral interventions, where students are forced to take notes in specific ways, have been found to be effective. Technology creates the opportunity to intervene on a larger number of note-taking behaviors. My research, however, indicates another compliance problem. As students can choose whether or not to take notes, **interventions students do not like can inhibit note-taking, which can have an adverse impact on learning outcomes.**

A careful design process produced an interface that restricted selection-size and was more satisfying than the unrestricted interface upon which it was based. Even though behavior was manipulated as desired, students did not perform better on

learning outcomes. In addition, the links between behavior and learning disappeared. It may be that the intervention was inappropriate for other reasons; allowing a student to click to select an entire sentence may have been as detrimental as large selections. Recommending behavioral changes while students were taking notes was also shown to change note-taking. This is particularly interesting given the failure of instruction or pre-training in the past. The in-context instruction facilitated by technology may allow for better note-taking instruction.

On the other hand **it may be better to use an analysis of note-taking behavior to update models of student knowledge**. According to the results described in this thesis, if a key idea is not recorded, or is recorded in a wordy fashion, it is more likely that the student does not understand that idea. The student can then be given targeted instruction based on how they have taken notes.

Students' Beliefs and Behaviors

This work also evaluated student beliefs regarding note-taking. In part, it confirmed results seen in previous studies. Students are conscious of both the processing and review benefits of note-taking, and are sensitive to interfaces they believe will interfere with either one. Interestingly, **students do not believe that the process of highlighting increases learning, while on the other hand this is their primary motivation for taking notes on separate documents.** On the other hand, they focus on the same content using both techniques. When highlighting or note-taking, they tend to focus on definitions, key terms, and key ideas. Highlighting also appeared to allow students with high levels of knowledge to skim the materials.

I also identified an interesting behavior that has previously been unreported in the literature. While reading on the computer, many students select text to help them read. They report several motivations for such **selection-to-read** behavior. It helps them attend to the reading task in general and focus on the specific content they are reading. Less frequently cited reasons include using selection to help read difficult type, and to serve as a bookmark. The behavior appears much like narrowing of focus readers of newspapers achieve through folding (Marshall & Bly, 2005).

A final interesting finding was that students often desired functionality that demonstrably reduces the efficiency of note-taking. Though students were faster when only allowed to copy-paste most wanted the ability to type, which slowed them down. While editing notes in a notepad slowed down the overall process without achieving improved learning outcomes, students wanted this ability as well.

Implications for Design

Efficiency vs. Desires

In this thesis, I have identified several situations in which there is a contrast between the efficiency of a note-taking interface and students' stated preference, where students prefer slower interfaces. The issues are as follows:

- **Typing**: When they are given a notepad, students want the ability to type and rate interfaces poorly if they do not provide such functionality. This is partly because they believe verbatim note-taking is detrimental to learning, though the results of these studies do not support such a belief. Students perform more quickly when they do not have the ability to type. While at first glance the design recommendation would be to develop user-friendly interfaces that only allow verbatim note-taking (such as highlighting), a limitation of these studies is that we could not estimate the value of using notes to connect the learning material with students' prior knowledge, as students were not observed to do so. As other research indicates that this may be a positive behavior, the design recommendation changes to developing interfaces that encourage verbatim notes when students are simply paraphrasing or rewording the learning material.
- **Organization**: Students given access to a notepad want the ability to reorganize those notes. However, editing notes significantly increases time on task without increasing learning outcomes. In particular, providing students with the ability to highlight without giving them access to a notepad actually improves satisfaction significantly, but students do not achieve review

benefits unless they are allowed to see the notepad as it is being constructed. Showing them the notepad creates the desire to organize notes, which slows them down without increasing learning.

In order to encourage a user-friendly efficient interface, two goals should be accomplished. First**, students should be discouraged from creating notes in their own words unless they are connecting the material with outside information**. Of course, the utility of these connections would also have to be explored experimentally. Secondly, **the interface should allow students to edit and reorganize their notes while at the same time discouraging them from doing so**. It may be appropriate to investigate automatic organization techniques that students find more useful than the basic ones implemented in this research. There are a variety of solutions that could be explored in design studies similar to the ones reported in this thesis. Dimensions could include how the notepad is presented (permanent/temporary, left/right, etc.), how typing is supported (always anchored in a highlight/freeform). This is only an example, and the design space would have to be defined in greater detail.

Designing for Learning

The design studies and their related experiments resulted in a set of guidelines for manipulating selection and for supporting highlighting. The full lists are available in chapters 7 and 9. I will use one to illustrate an important point. The initial design study concluded that the interface should not change a selection once it is made. This may appear to be an obvious recommendation for most user interface designers, as changing selections breaks basic usability heuristics. However, learning interfaces are often intentionally difficult in order to motivate learning. The easiest interface solves the problem for the user. Unfortunately, the user is unlikely to learn much from such an interface.

It is therefore important to distinguish appropriate frustrations from inappropriate frustrations. Changing a user's selections is inappropriate because it interferes with behaviors that have nothing to do with the target of note-taking. In this case, the

design study found that people who selected to read found this very frustrating. Optional educational interfaces have the additional challenge of reducing the overall degree of frustration to a level which will not reduce adoption, as students can simply avoid using interfaces they do not like.

Design Process

This work includes two examples of a mixed-methods iterative design process. The process involves describing a design space with regards to interface dimensions and testing different combinations of those dimensions with users, who are asked to review the interfaces as well as the dimensions. This testing is done iteratively so that hypotheses regarding the effect of specific dimension combinations can be tested. Though it does not produce empirically validated conclusions, it does provide behavioral and attitudinal evidence for design decisions.

This process produced effective solutions to two different types of design problems. In the first study, I developed an interface that encouraged students to take shorter notes, but which actually increased user satisfaction relative to an unrestricted tool. The second study produced a highlighting tool that effectively navigated divisive interface dimensions. Though the final interaction was non-standard, it earned positive user ratings, and no student expressed dislike for the interface.

It would be interesting to explore the boundaries of this design process. It is clear that this will only work for design spaces that can be described with regards to interface dimensions. This process would be inappropriate for spaces in which the cost of building a single interface is high, or for which each dimension requires a significant amount of development work to implement. It may also require design problems for which the interface has specific goals. In the first study the interface was allowing students to select text. In the second it was allowing students to highlight text. More nebulous problems, such as helping a person find some arbitrary bit of information, may not be so easily definable.

Limitations

There are several limitations of this work that may have affected the learning outcomes observed in this work. First, all experiments followed the same testing paradigm, which was based on the prior literature. In these studies the immediate post-test, which is one of the more common tests in the literature, may have served as a sort of delayed note-taking. Immediately after reading the learning material, students were asked to define the very key ideas around which the tests were based. This may have been a form of structured note-taking. This form of note-taking, which identifies key ideas for students and is described in greater detail in chapter one, can be an effective form of note-taking. Processing differences between interfaces may thus have been washed out by having all participants define the key ideas immediately after reading. Future experimentation should test this hypothesis by doing only delayed testing.

The results may also be content specific. All studies used the same materials, in part because the learning content and tests were built around a set of key ideas that facilitated analyses connecting behavior and learning. In addition, other materials that showed different learning results for different note-taking treatments were unavailable or had not included a no-notes treatment. The materials used in this research are difficult and unfamiliar to most students of the population tested. This may also be why students were not observed to bring outside information into their notes, as they may not have had easily relatable knowledge. In future studies, it may be more appropriate to use content with a larger variety of tested ideas.

References

1. Abowd, G.D. (1999) Classroom 2000: An experiment with the instrumentation of a living educational environment. *IBM Systems Journal, 38*(4), 508-530.
2. Adler, M.J., van Doren, C. (1972). *How to Read a Book.* New York: Simon and Schuster
3. Anderson, R.J., Hoyer, C., Wolfman, S.A., Anderson, R. A Study of Digital Ink in Lecture Presentation. (2004). In Proceedings of *ACM Conference on Human Factors in Computing Systems (CHI '04).* pp.567-574 ACM Press.
4. Annis, L., Davis., J.K. (1975). The effect of encoding and an external memory device on notetaking, *Journal of Experimental Education, 44*(2), 74-76
5. Annis, L., Davis., J.K. (1978). Study Techniques and Cognitive Style: Their Effect on Recall and Recognition, *Journal of Educational Research, 71*(3), 175-178
6. Armel, D. (1995). Something New about Notetaking: A computer-based instructional experiment, In proceedings of *Association of Small Computer Users in Education Summer Conference.* pp. 22-29
7. Baker, R.S., Corbett, A.T., Koedinger, K.R., Wagner, A.Z. (2004). Off-Task Behavior in the Cognitive Tutor Classroom: When Students "Game The System". In Proceedings of *ACM Conference on Human Factors in Computing Systems (CHI '04).* pp.383-390. ACM Press
8. Bauer, A., and Koedinger, K. "Developing a Note Taking Tool from the Ground Up". In Proceedings of *Ed-Media 2005.* pp. 4181-4186 AACE Press
9. Bauer, A., Koedinger, K.R. Pasting and Encoding: Note-taking in Online Courses. In Proceedings of *IEEE International Conference on Advanced Learning Technologies (ICALT '06).* pp. 789-793 IEEE Press
10. Bauer, A., Koedinger, K.R. (2007) Selection-Based Note-Taking Applications. In Proceedings of *ACM Conference on Human Factors in Computing Systems (CHI '07).* pp.981-990 ACM Press
11. Bessler, S., Hager, M., Benz, H., Mecklenburg, R. Fischer, F. (1997). DIANE: A Multimedia Annotation System, Proceedings of the Second In Proceedings of *European Conference on Multimedia Applications, Services and Techniques* pp.183-198
12. Boren, T., Ramey, J. (2000). Thinking aloud: reconciling theory and practice, *IEEE Transactions on Professional Communication, 43*(3), 261-278
13. Bretzing, B. H., & Kulhavy, R. W. (1979) Note-taking and depth of processing. *Contemporary Educational Psychology 4*, 145-153.
14. Bretzing, B.H., Kulhavy, R.W. (1981). Note-taking and Passage Style. *Journal of Educational Psychology, 73*(2), 242-250
15. Brush, A.J.B., Bargeron, D., Grudin, J., Borning, A., and Gupta, A. (2002). Supporting Interaction Outside of Class: Anchored Discussions vs Discussion Boards. In Proceedings of *ACM Conference on Computer Supported Collaborative Learning.* (CSCL '02). 425-434. ACM Press
16. Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In *Metacognition: Knowing about knowing.* (pp.185-205). MIT Press.

References

17. Card, S.K., Moran, T., & Newell, A. (1983). *The psychology of human-computer interaction.* Hillsdale, NJ: Lawrence Erlbaum Associates

18. Carrier, C.A, Williams, M.D., Dalgaard, B.R. (1988). College Students' Perceptions of Notetaking and their Relationship to Selected Learner Characteristics and Course Achievement. *Research in Higher Education, 28*(3), 223-239

19. Carter, John F., Van Matre, Nicholas H. (1975) Note taking versus note having, *Journal of Educational Psychology, 67*(6), 1975, pp. 900-904.

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

21. 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.

22. Crawford, C.C. (1925) The correlation between college lecture notes and quiz papers. *Journal of Educational Research, 12*(4), 282-291.

23. Davis, R.C. Landay, J.A. Chen, V. Huang, J. Lee, R.B., Li, J.L., Morrey, C.B. Schleimer, B. Price, M.N Schilit, B.N. (1999) NotePals: Lightweight note sharing by the group, for the group. In Proceedings of *ACM Conference on Human Factors in Computing Systems (CHI '99)*, 338-345, ACM Press

24. Denoue, L., Vignollet. L. (2000) An annotation tool for web browsers and its applications to information retrieval. In Proceedings of *RIAO2000*

25. Dyer, J.W., Riley, J.,Yekovich, F.R. (1979) An analysis of three study skills: Notetaking, summarizing, and rereading. *Journal of Educational Research 73*, 3-7.

26. Fisher, J.L., Harris, M.B. (1973). Effect of Note-taking and Review on Recall. *Journal of Educational Psychology 65*(2), 321-325

27. Fowler, R.L., Barker, A.A. (1974) Effectiveness of Highlighting For Retention of Text Material. *Journal of Applied Psychology, 59*(3), 358-364

28. Fox, B.J., Siedow, M.D. (1985) An Investigation of the Effects of Note Taking on College Students' Recall of Signalled and Unsignalled Text, *Journal of Research and Development in Education, 18*(3), 29-36

29. Gilmore, D.J. (1996). The Relevance of HCI Guidelines for Educational Interfaces. *Machine-Mediated Learning 5*(2) 119-133

30. Golightly, D., Gilmore, D. (1996). Harnessing the Interface for Domain Learning. In the Proceedings of the *Conference companion on Human factors in computing systems: common ground*, 37-38

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

32. Howe, M.J. (1970). Using students' notes to examine the role of the individual learner in acquiring meaningful subject matter. *Journal of Educational Research 64*(2), 61-63

33. Igo, B.L., Bruning, R., McCrudden, M.T. (2005). Exploring Differences in Students Copy-and-Paste Decision Making and Processing: A Mixed Methods Study. *Journal of Educational Psychology 97*(1), 103-116

34. Johnson, L.L. (1988). Effects of Underlining Textbook Sentences on Passage and Sentence Retention. *Reading and Research Instruction 28*(1), 18-32

References

35. Jung, B., Yoon, I., Lim, H., Ramirez-Weber, F. A., Petkovic, D. (2006) Annotizer: User-friendly WWW Annotation System for Collaboration in Research and Education Environments, In Proceedings of the *IASTED International Conference on Web Technologies, Applications and Services, WTAS 2006*

36. Kam, M., Wang, J., Iles, A., Tse, E., Chiu, J., Glaser, D., Tarshish, O., and Canny, J.F. (2005). Livenotes: a system for cooperative and augmented note-taking in lectures. In Proceedings of *ACM Conference on Human Factors in Computing Systems (CHI '05)*. pp.531-540. ACM Press

37. Katayama, A.D., Crooks, S.M. (2001) Online Notes: Differential Effects of Studying Complete or Partial Graphically Organized Notes. *The Journal of Experimental Education, 71*(4), 293-312

38. Katayama, A.D., Shambaugh, R.N., Doctor, T., (2005). Promoting Knowledge Transfer with Electronic Note Taking, *Teaching of Psychology, 32*(2), 129-131

39. Kiewra, K.A., Benton, S.L., Kim, S.I., Risch, N., Christensen, M. (1995) Effects of Note-Taking Format and Study Technique on Recall and Relational Performance., *Contemporary Educational Psychology, 20*, 172-187

40. Kiewra, K.A., DuBois, N.F. (1991). Note-taking Functions and Techniques. *Journal of Educational Psychology 83*(2) 240-245

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

42. Kiewra, K.A., Fletcher, H.J., (1984) The Relationship Between Levels of Note-Taking and Achievement., *Human learning, 3*, 273-180

43. Kiewra, K.A. (1985). Investigating Notetaking and Review: A Depth of Processing Alternative. *Educational Psychologist, 20*(1), 23-32

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

45. Kintsch, Walter. (1986). Learning from Text. *Cognition and Instruction 3*, 87 - 108.

46. Kintsch, W. (1994), Text Comprehension, Memory, and Learning *American Psychologist, 49*(4), 294-303

47. Kintsch, W. (1998) *Comprehension: A paradigm for cognition.* New York: Cambridge University Press.

48. Kobayashi, K. (2005). What Limits the Encoding Effect of Note-Taking? A meta-analytic examination., *Contemporary Educational Psychology, 30*, 242-262

49. 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

50. Kulhavey, R. W., Dyer, J. W., and Silver, L. (1975) The effects of note-taking and test-expectancy on the learning of text material. *Journal of Educational Research, 68*, 363-365.

51. Kurhila, J.; Miettinen, M.; Nokelainen, P.; Tirri, H. (2004). The role of the learning platform in student-centred e-learning" *IEEE ICALT'04*, pp.540 - 544

References

52. 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.* , 2737-2742

53. Lebow, D., Lick, D., Hartman, H., (2004) Interactive Annotation for Teaching and Learning. *Society for Information Technology and Teacher Education International Conference,* pp.1781-1786

54. Lewis, C., Brand, C., Cherry, G., Rader, C. (1998), Adapting User Interface Design Methods to the Design of Educational Activities. In Proceedings of *ACM Conference on Human Factors in Computing Systems (CHI '98)*, 619-626

55. Lonka, K., Lindblom-Ylanne, S., Maury, S. (1994) The Effect of Study Strategy on Learning From Text. *Learning and Instruction, 4*, 253-271

56. Mannes, S. M., & Kintsch, W. (1987). Knowledge organization and text organization. *Cognition and Instruction, 4*(2), 91–115.

57. Marshall, C.C. (1998). Towards an Ecology of Hypertext Navigation. IN Proceedings of the ACM Conference on Hypertext (*Hypertext '98)* pp.40-49. ACM Press

58. Marshall, C.C. & Bernheim Brush, A.J. (2004) Exploring the Relationship between Personal and Public Annotations. In Proceedings of *Digital Libraries (DL '04)*, ACM Press, 349-357

59. 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* (JCDL '05), pp. 225-234. ACM Press

60. Mason, H., Rebelsky, S., & Luebke, S. (1999). Annotating the World-Wide Web. *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, 409-414. AACE Press

61. McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*, 1-43

62. 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

63. Nielsen, J. (1993). Iterative User-Interface Design, *IEEE Computer 26*(11) pp.32-41

64. Nye, P., Crooks, T., Powley, M., & Tripp, G. (1984). Student note-taking related to university examination performance. *Higher Education 13*, 85-97.

65. Ovsiannikov, I. A., Arbib, M. A., and McNeill, T. H. (1999). Annotation technology. *International. Journal of. Human-Computer Studies* 50(4), 329-362.

66. Palmatier, R.A., Bennett, J.M., (1974) Notetaking Habits of College Students, *Journal of Reading, 18*, 215-218

67. Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003).Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63-71.

68. Pauk, Walter *How to study in college.* 7th ed. Boston: Houghton Mifflin, 2000.

69. Peper, R.J., Mayer, R.E. (1986). Generative Effects of Note taking During Science Lectures. *Journal of Educational Psychology, 78*(1), 34-38

70. Peters, D.L. (1972) Effects of note taking and rate of presentation on short-term objective test performance. *Journal of Educational Psychology, 63*, 276-280

References

71. Peverly, S.T., Brobst, K.E., Graham, M., Shaw, R. (2003). College Adults Are Not Good at Self-Regulation: A Study on the Relationships of Self-Regulation, Note Taking, and Test Taking. *Journal of Educational Psychology, 95*(2), 335-346

72. Quade. A. M. (1995) An assessment of retention and depth of processing associated with notetaking using traditional paper and pencil and on-line notepad during computer-delivered instruction. *AECT'95*

73. Reder, L.M. (1985). Techniques available to author, teacher and reader to improve retention of main ideas of a chapter. In S. Chipman, J. Segal, & R. Glazer (Eds.), *Thinking and learning skills: Current research and open questions, Vol. 2.* Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 37-64.

74. Rice, G.E. (1994). Examining Constructs in Reading Comprehension Using Two Presentation Modes: Paper vs. Computer, *Journal of Educational Computing Research, 11*(2), 153-178

75. Rickards, J.P., August, G.J. (1975) Generative Underlining Strategies in Prose Recall. *Journal of Educational Psychology, 67*(6), 860-865

76. Rickards, J.P., Friedman, F. (1978) The Encoding Versus the External Storage Hypothesis in Note Taking. *Contemporary Educational Psychology, 3*, 136-143

77. Rickards, J.P., Fajen, B.R., Sullivan, J.F., Gillespie, G. (1997). Signaling, Notetaking and Field Independence-Dependence in Text Comprehension and Recall. *Journal of Educational Psychology, 89*(3), 508-517

78. Robinson: D.H., Kiewra, K.A. (1995). Visual Argument: Graphic Organizers are Superior to Outlines in Improving Learning From Text, *Journal of Educational Psychology, 87*(3) 455-467

79. Santa, C.M., Abrams, L., Santa, J.L (1979) Effects of Notetaking and Studying on the Retention of Prose. *Journal of Reading Behavior, 11*(3) 247-260

80. Scheines , R., Leinhardt, G., Smith, J., Cho, K. (2005) Replacing lecture with web-based course materials. *Journal of Educational Computing Research, 32* (1), 1-25.

81. Schellings, G.L.M., Van Hout-Wolters, B.H.A.M. (1995). Main Points in an Instructional Text, as Identified by Students and by Their Teachers, *Reading Research Quarterly, 30*(4), 742-756

82. Schilit, B.N. Price, M.N. Golovchinsky, G. (1998). Digital library information appliances. In Proceedings of the *ACM Conference on Digital Libraries (DL '98)*, 217-226, ACM Press

83. Schultz, C.B., Di Vesta, F.J. (1972). Effects of Passage Organization and Note Taking on the Selection of Clustering Strategies and On Recall of Textual Materials. *Journal of Educational Psychology, 1972 63*(3), 244-252

84. Shimmerlik, S.M., Nolan, J.D. (1976) Reorganization and the Recall of Prose. *Journal of Educational Psychology 68*(6), 779-86

85. Sweller, J., Van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-296.

86. Truong, K.N., Abowd, G.D., and Brotherton, J.A. (1999) Personalizing the Capture of Public Experiences. In Proceedings of *ACM Conference on User Interface Software and Technology. (UIST'99).* 121-130. ACM Press

References

87. Van Meter, P.M., Yokoi, L., Pressley, M. (1994). College Students' Theory of Note-taking Derived From Their Perceptions of Note-taking. *Journal of Educational Psychology 86*(3). 323-338

88. Van Oostendorp, H., (1996) Studying and Annotating Electronic Text. In J.F. Rouet, J. Levonen, A. Dillon & R.J. Spiro (Eds.), *Hypertext and Cognition.* Hillsdale, NJ: Lawrence Erlbaum Associates.

89. Wang, C.Y., Wei, F.H., Chao, P.Y., Chen, G.D. "Extending e-books with contextual knowledge recommenders by analyzing personal portfolio and annotation to help learners solve problems in time." In Proceedings of the *International Conference on Advanced Learning Technologies. (*ICALT"04) pp. 306 - 310

90. Wang, CY & Chen, G.W. (2004) Extending E-books with Annotation, Online Support and Assessment Mechanisms to Increase Efficiency of Learning. In Proceedings of *SigCSE 2004*. 132-136. ACM Press

91. Ward, D.J., Blackwell, A.F., MacKay, D.J.C. (2000). Dasher - a data entry interface using continuous gestures and language models. *UIST 2000.* pp.129-137

92. Wei, F.H., Lee, L.Y., Chen, G.D. (2004). Supporting adaptive mentor by student preference within context of problem-solving learning" IEEE In Proceedings of the *International Conference on Advanced Learning Technologies (ICALT"04).* pp.789 - 791

93. Wilcox, L.D., Schilit, B.N., Sawhney, N.N.(1997), Dynomite: A Dynamically Organized Ink and Audio Notebook, In Proceedings of *ACM Conference on Human Factors in Computing Systems (CHI '97)* pp. 186-193 ACM Press

94. Wilkerson, M., Griswold, W. and Simon, B. (2005). Ubiquitous Presenter: Increasing Student Access and Control in a Digital Lecturing Environment. In Proceedings of *ACM Conference on Human Factors in Computing Systems (CHI '05).* pp.116-120. ACM Press

95. Wiley, J. (2001) Supporting understanding through task and browser design. *Proceedings of the Twenty-third annual Conference of the Cognitive Science Society1* pp. 1136-1143

96. Williams, R.L., & Eggert, A.C. (2002). Notetaking in college classes: Student patterns and instructional strategies. *The Journal of General Education 51*(3), 173-199

97. Wolfe, J. (2002). Annotation Technologies: A Software and Research Review. *Computers & Composition 19*, 471-491

98. Wolfe, J. L., and Neuwirth, C. M. (2001). From the margins to the center: The future of annotations. *Journal of Business and Technical Communication 15*(3), 333-371.

99. Wright P (1988) The need for theories of NOT reading: some psychological aspects of the human-computer interface. In B AG. Elsendoorn and H Bouma (eds) *Working Models of Human Perception.* London: Academic Press. p319-340

Appendix A: Key Idea Definitions

⇒*The studies in this thesis used materials from an online course in Causal and Statistical Reasoning. The course can be viewed at http://www.cmu.edu/oli. The version used in this thesis is from January 2004. Each module in this course is built around a set of key ideas defined by the instructor. The following are definitions of each key idea taken from the learning materials used in the final four studies. Though not all forms are in this appendix, the following cover all definition forms.*

Causal Relativity

1. *General*: Causal Relativity involves the idea that direct causation is not an absolute concept, but one that is relative, at least to the background conditions and to the set of variables under discussion.
2. *Set of Variables*: Claims about the direct causal relationships between one variable X and another Y must also be judged relative to the set of variables Z that are explicitly under consideration.
   a. We call the set of variables explicitly under consideration the causal system.
3. *Background Conditions*: Although the background conditions are typically not even mentioned, causal claims take on meaning only against some set of background conditions.

Causal Assignment

1. A causal assignment is one particular assignment of values to all the variables in a given causal system except one variable designated as the effect.
2. Intervening on all the variables in a system besides an effect Y is to make a causal assignment.

Intervention

1. Causal Assignments are Interventions, not Observations
2. To learn which causal assignment(s) produces malaria, we must be able to intervene and produce each assignment.
3. If, for example, we are in an area we can only observe things and not intervene to control them, then according to our definition we cannot directly decide on causal questions in that domain.

Causal Assignment: Number

1. In general, if there are N variables, then the number of causal assignments is equal to: (# of values for Var. 1) X (# of values for Var. 2) X ... X (# of values for Var. N)

Test Pair

1. Two causal assignments C1 and C2 are a test pair of causal assignments for X if and only if they are identical except for the values assigned to variable X

Response Structure

1. The route to a general account of causation among variables goes through tables which include all the possible causal assignments for some effect, as well as some description of the effect in each such assignment. We call such tables response structures.
2. One way to reliably isolate which among many potential causes of Y is a real cause is to intervene on all the variables in a system besides Y and give them every possible value.

Direct Causation

1. In a system of variables S, X is a direct cause of Y relative to S if and only if there is at least one test pair of causal assignments for X in the response structure for Y across which Y differs.
2. Put in the language of causal assignments and response structures: if there are any two causal assignments that are identical except for the value assigned to X, and there is a difference in the effect Y, then X is a direct cause of Y.
3. If we can wiggle X while we hold everything else constant and produce a change in Y (the test pair of assignments for X), then X is a direct cause of Y.
4. The idea, first made clear by the 18th Century philosopher John Stuart Mill, is to hold everything else in the system constant, and only vary the potential cause. If X causes Y, then intervening to change X should change Y in some way.

Interaction

1. When the influence of one direct cause depends upon the state of another direct cause, then we say the causes interact.
2. Two causes are "interacting causes" when the influence of one cause depends on the value of the other cause.

Response Structure Uniformity

1. A population has response structure uniformity for a given effect if every individual in the population has the same response structure for that effect.
2. … response structure uniformity: every individual in the population is governed by the same response structure.

Appendix B: Examples from Module

⇒*To give a sense of how the learning content used is formed, two examples from the material are given.. The headers are from the module. The first contains both a statement that causal generalizations are relative to background conditions and an example of why this is so:*

Background Conditions

Causal generalizations are always to be understood as *relative to a particular set of background conditions*. For example, consider the claim: "Eating red meat several times a week reduces the length of one's life." In America in the beginning of the 3rd millennium, where diets tend to contain lots of calories and lots of calories from saturated fat, we might accept this causal claim as true. In what is now Northern Arizona around 1500, however, eating red meat more than once a week would almost certainly increase life span, mostly because the normal diet was so spare that groups in some cases resorted to Cannibalism. What's the difference? The background conditions, which in this case include the "normal" diet for the time and place. Although the background conditions are typically not even mentioned, causal claims take on meaning only against some set of background conditions.

⇒*The following is an example of why causal generalizations are relative to the set of variables that are included in the causal system being studied:*

The Set of Variables

Claims about the direct causal relationships between one variable X and another Y must also be judged *relative to the set of variables Z that are explicitly under consideration.*

Example: Causal Relativity: Lighting Matches

For example, consider four variables about matches: Match Color [blue, red], Match Struck [yes, no], Match Tip Temperature [above 300, below 300], and Match Lit [yes, no]. Asked whether striking a match is a direct cause of the match lighting, that is, whether the variable Match Struck is a direct cause of the variable Match Lit, the answer depends on the set of variables considered.

If the set is either:

* {Match Color, Match Struck, Match Lit}, or
* {Match Struck, Match Lit}

then the answer is yes. If the set includes Match Tip Temperature, however:

* {Match Color, Match Struck, Match Tip Temperature, Match Lit}, or
* {Match Struck, Match Tip Temperature, Match Lit}

then the answer is no. In these sets, Match Struck is only an indirect cause of Match Lit. Match Struck is a direct cause of Match Tip Temperature, which in turn is a direct cause of Match Lit.

⇒*Finally, here is one of several examples explaining interacting causation:*

Example: Non-Interacting Causes: the Battery, Switch, and Light Bulb

Consider the Battery, Switch, and Light Bulb case in slightly more detail. The response structure for the light bulb is as follows.

Response Structure for the Light Bulb:

Assignment	*Battery*	*Switch*	Effect: *Garage Light*
1	Charged	Closed	On
2	Charged	Open	Off
3	Dead	Closed	Off
4	Dead	Open	Off

Across one test pair for the switch: assignments 1 and 2, when the switch 2 is set to closed, the light is on and when the switch is set to open, the light is off. Across the other test pair for the switch, assignments 3 and 4, the light is off no matter whether the switch is set to opne or closed. The *influence* of the switch *depends* on the state of the battery.

Appendix C: Quizzes from Final Study

The following are the quiz items from the final study. Most of these items were used throughout all of the studies in this thesis. Others are new versions created when items with low discriminability were thrown out. All are based on the 12-item multiple-choice test initially created by the course developer. The related key idea is in italics next to the question statement. See Appendix X for answers to the definition questions.

Quiz A

1. Consider the following causal system: {Tired[yes, no], On Couch[yes, no], Watched Scary Movie[yes, no], Sleeps[Yes, No]}, where Sleeps is the effect. Please draw a response structure diagram where only On Couch and Watched Scary Movie are direct causes of Sleeps.
Direct Causation- Multiple possible answers, one follows:

Tired	On Couch	Watched Scary	Sleeps
Yes	Yes	Yes	Yes
Yes	Yes	No	No
Yes	No	Yes	Yes
Yes	No	No	Yes
No	Yes	Yes	Yes
No	Yes	No	No
No	No	Yes	Yes
No	No	No	Yes

2. According to the module, claims about direct causal relationships between variables are relative to what two major things?
Causal Relativity

      a._**Causal system (set of variables)**_____
      b._**Background conditions**_____

3. Consider a causal system involving: {Stick thrown[yes, no] and Stick retrieved[yes, no]}, where Stick Retrieved is treated as the effect. Which of the following populations is LEAST likely to have response structure uniformity in this case?
Response Structure Uniformity

A. A group of cats and dogs
B. A group of border collies [a breed of dog]
C. A group of dogs.
D. A group of cats.

4. If variable X interacts with variable Y to produce an effect Z, then which ONE of the following MUST be true?
Interaction

A. Variable X cannot be a direct cause of the effect Z
B. Variable X is a direct cause of the effect Z
C. For all test pairs for X, there is a difference in the effect Z
D. Variable X is an indirect cause of the effect Z

5. How is direct causation defined in the material you have studied in this experiment? Give the definition with regards to test pairs, causal assignments, or the values of variables.
Direct Causation

6. Imagine that the causal system consists of the variables {Drinks [yes, no], Stays out Late [yes, no], Grades [good, poor]}. You are interested in the influence of the variables Drinks and Stays out Late on Grades. You take your friend Herb to a local bar the night before an exam, making sure he stays out late. While at the bar, you carefully record what he drinks. This is not a causal assignment, because:
Choose exactly one of the following:
Intervention

A. A value has not been assigned to the effect, Grades
B. There are other causes of Grades, like intelligence, and you have not taken these into account
C. You have only observed the values of Drinks and Stays out Late, not assigned them
D. You only assigned a value to Stays out Late, not Drinks

7. If two people are in the same causal assignment, which ONE of the following is true?
Causal Assignment

A. They must have the same response structure
B. They will exhibit the same value for the effect
C. They have the same value for all potential causes
D. They differ with regards to only one potential cause

8. Consider the following causal system: {bright light[yes, no], puff of air[yes, no], blinks[yes, no]} where blinks is the effect. Please draw a response structure diagram for which bright light and puff of air interactively cause blinks.
Interaction- Multiple possible answers, one follows:

Bright Light	Puff of Air	Blinks
Yes	Yes	Yes
Yes	No	No
No	Yes	No
No	No	No

9. Consider the causal system {Parks Outside[yes, no], Friendly[Yes, No]}. Suppose that you are interested in the influence of parking outside on friendliness. You observe your neighbors Jim and Larry, and you carefully record your observations in the following table:

	Parks Outside	Friendly
Jim	No	Yes
Larry	Yes	No

This is not a response structure for Friendliness (i.e. for which friendliness is the effect) because:

Choose exactly one of the following:
Intervention

A. Not everyone who parks outside is unfriendly
B. Not enough variables have been included
C. You have observed the value of friendly, not assigned it
D. You have observed the value of parks outside, not assigned it

10. Consider the following causal system: {bright light[yes, no], puff of air[yes, no], blinks[yes, no]} where blinks is the effect. Please draw a response structure diagram in which only puff of air is a direct cause of blinks.
Direct Causation- Multiple possible answers, one follows:

Bright Light	Puff of Air	Blinks
Yes	Yes	Yes
Yes	No	No
No	Yes	Yes
No	No	No

11. Homenet researchers found that greater levels of internet use were linked with greater levels of depression in 1995, but linked with lower levels of depression in 2000. They believe that in 2000 more people were online talking to friends, whereas in 1995 people were more likely to be online talking to strangers. This is most clearly an example of differences in which ONE of the following: *Causal Relativity-Background Conditions*

A. The Causal System
B. Causal Assignment
C. The Background Conditions
D. Response Structure

12. When do two potential causes interact?
Interaction

13. With respect to which of the following causal systems is exposure to a flu virus a direct cause of getting the flu?
Choose exactly one of the following:
Causal Relativity-Set of Variables

A. Exposed to the Flu Virus [yes, no], Infected with the Flu Virus [yes, no], Gets the Flu [yes, no]
B. Exposed to the Flu Virus [yes, no], Gets the Flu [yes, no], Takes Medication for Temporary Relief of Flu Symptoms [yes, no]
C. Exposed to the Flu Virus [yes, no], Infected with the Flu Virus [yes, no], Gets the Flu [yes, no], Sleeps at Least 8 Hours per Night [yes, no]
D. Eats Citrus Fruit [yes, no], Exposed to the Flu Virus [yes, no], Infected with the Flu Virus [yes, no], Gets the Flu [yes, no]

14. Given a causal system with 3 potential causes, <u>one</u> of which has 2 possible values [Yes, No] and <u>two</u> of which have 3 possible values [Good, Average, and Poor], how many causal assignments will be in a response structure for the causal system?
Causal Assignments-Number

18

15. What is a test pair?
Test Pairs

16. Most men who do not shave for an extended period of time will grow a beard. Most women who do not shave for the same amount of time will not grow a beard. This is a difference with regards to which of the following:
Causal Assignments vs. Response Structures

A. Causal Assignments
B. Response Structures
C. Test Pairs
D. Interactions

17. Consider the causal system : {Sneezes[yes, no], Blinks[yes, no], Turned Away[yes, no], and Sees Shooting Star [yes, no]} in which Sees Shooting Star is the effect. Give an example of a test pair for Sneezes.
Test Pair- *Multiple possible answers, one follows:*

Sneezes	Blinks	Turned Away	Sees Shooting Star
Yes	Yes	Yes	?
No	Yes	Yes	?

18. Which of the following is most clearly not a background condition to which the causal generalization, "Turning the key in the ignition causes the car to start," is relative?
Choose exactly one of the following:
Causal Relativity- Background Conditions

A. Battery Charged
B. Pedal Depressed
C. Gas Tank Empty
D. Starter is functional

19. What do we mean when we say a population has Response Structure Uniformity?
Response Structure Uniformity

20. Consider the causal system involving: {Exposed [yes, no], Innoculated [yes, no], Previously Infected [yes, no], and Gets Chicken Pox Rash [yes, no], and the following Response Structure for Gets Chicken Pox Rash.
Test Pair

Assignment	Exposed	Innoculated	Previously Infected	Chicken Pox
1	Yes	Yes	Yes	No
2	Yes	Yes	No	No
3	Yes	No	Yes	No
4	Yes	No	No	Yes
5	No	Yes	Yes	No
6	No	Yes	No	No
7	No	No	Yes	No
8	No	No	No	No

Which of the following are test pairs for Innoculated?
Choose exactly one of the following:
Test Pairs

A. 6 and 8
B. 1 and 2
C. 4 and 5
D. 3 and 6

21. Consider this response structure:

Causal Assignment	Acupuncture Therapy	Counseling	Breaks Addiction
1	Yes	Yes	Yes
2	Yes	No	No
3	No	Yes	No
4	No	No	No

Which of the following are test pairs for Counseling?
Choose exactly one of the following:
Test Pairs

A. 3 and 4
B. 1 and 3
C. 2 and 4
D. 2 and 3

22. What is a response structure? Give the definition with regards to test pairs, causal assignments, or the values of variables.
Response Structure

23. How is a causal assignment created?
Causal Assignment

Quiz B

1. What do we mean when we say a population has Response Structure Uniformity?
Response Structure Uniformity

2. Consider the following response structure:

Causal Assignment	Got Drunk	Studied	Passed Exam
1	Yes	Yes	No
2	Yes	No	No
3	No	Yes	Yes
4	No	No	No

Which of the following are test pairs for Studied?
Choose exactly one of the following:
Test Pairs

A. 1 and 2
B. 2 and 3
C. 1 and 3
D. 2 and 4

3. Consider the causal system {Works >40 hour weeks [yes, no], happy[yes, no]}. Suppose that you are interested in the influence of working more than 40 hour weeks on being happy. You ask your friends Sally and Amber how many hours they work, and whether they are happy, recording their responses in the following table.

	Works > 40 hours/week	Happy
Sally	No	Yes
Amber	Yes	No

This is not a response structure for happiness (i.e. for which Happy is the effect) because:
Choose exactly one of the following:
Intervention

A. Not everyone who works more than 40 hour weeks is unhappy
B. You have observed the value of Happy, not assigned it
C. You have observed the value of works > 40 hours/week, not assigned it
D. Not enough variables have been included

4. When do two potential causes interact?
Interaction

5. If you take two people with allergies to cats and bring one to a room where there is a cat and one in a room without cats, the first person will have an allergic reaction while the other will not. This is most clearly due to a difference in which of the following:
Causal Assignments and Response Structures

A. Response Structures
B. Causal Assignments
C. Interactions
D. Test Pairs

6. How is a causal assignment created?
Causal Assignment

7. Consider the causal system involving: {Gas Tank Empty [yes, no], Key Turned [yes, no], Starter Engaged [yes, no], and Car Starts [yes, no]}, and the following Response Structure for Car Starts.

Assignment	Gas Tank Empty	Key Turned	Starter Engaged	Car Starts
1	Yes	Yes	Yes	Yes
2	Yes	Yes	No	No
3	Yes	No	Yes	No
4	Yes	No	No	No
5	No	Yes	Yes	No
6	No	Yes	No	No
7	No	No	Yes	No
8	No	No	No	No

Which of the following are test pairs for Key Turned?
Choose exactly one of the following:
Test Pairs

A. 1 and 5
B. 2 and 4
C. 3 and 6
D. 6 and 7

8. How is direct causation defined in the material you have studied in this experiment? Give the definition with regards to test pairs, causal assignments, or the values of variables.
Direct Causation

9. What is a test pair?
Test Pair

10. Which of the following is a background condition to which the causal claim, "Pushing the button on my desktop computer causes it to turn on," is relative?
Choose exactly one of the following:
Causal Relativity-Background Conditions

A. The computer is plugged in
B. The computer is black with blue trim
C. The monitor is already on
D. The monitor is connected to the computer

11. Consider the following causal system: {Uses Pesticide[yes, no], Rains[yes, no], Plants Grow[yes, no]} where Plants Grow is the effect. Please draw a response structure diagram in which only Uses Pesticide is a direct cause of Plants Grow.
Direct Causation-Multiple possible answers, one follows:

Uses Pesticide	Rains	Plants Grow
Yes	Yes	Yes
Yes	No	Yes
No	Yes	No
No	No	No

12. With respect to which of the following causal systems is pressing down on the brake pedal a direct cause of the car coming to a stop?
Choose exactly one of the following:
Causal Relativity-System of Variables

A. Press down on pedal[yes, no], headlights on[yes, no], brakes engage[yes, no], car stops[yes, no]
B. Press down on pedal[yes, no], headlights on[yes, no], car stops[yes, no]
C. Press down on pedal[yes, no], headlights on[yes, no]
D. Press down on pedal[yes, no], brakes engage[yes, no], car stops[yes, no]

13. Given a causal system with 3 potential causes, <u>one</u> of which has 3 possible values [Good, Average and Poor] while the other <u>two</u> have 2 possible values["Yes" and "No"], how many causal assignments will be in a response structure for the causal system.
Causal Assignments-number
   12

14. If variable X interacts with variable Y to produce an effect Z, then which ONE of the following CANNOT be true?
Interaction

A. There are no test pairs for X for which the effect Z differs.
B. Variable Y is a direct cause of variable Z
C. Variable X is a direct cause of variable Z
D. There is another variable which directly causes Z

15. In Arizona in the 1500s, increased consumption of red meat was associated with longer lifespans, whereas today increases in consumption of red meat may be linked with shorter lifespans. We believe that this is due to differences in the normal diet for the times. If true, this is most clearly an example of differences in which ONE of the following:
Causal Relativity-Background Conditions

A. Potential Causes
B. The Background Conditions
C. The Causal System
D. Response structure

16. Consider the following causal system: {exercised[yes, no], sleep deprived[yes, no], and tired[yes, no]}, where tired is the effect. Please draw a response structure diagram in which exercised and sleep deprived interactively cause tired.
Interaction-Multiple possible answers, one follows:

Exercised	Sleep Deprived	Tired
Yes	Yes	Yes
Yes	No	No
No	Yes	No
No	No	No

17. According to the module, claims about direct causal relationships between variables are relative to what two major things?

      a._**Causal system (set of variables)**_____

      b._**Background conditions**_____

18. Consider the causal system {Sober [yes, no], Dances [Yes, No]}. Suppose you want to know whether Sobriety influences Dancing. You go to a school party, and observe the following:

	Sober	Dances
Jenny	Yes	Yes
Sara	No	No

This is not a response structure for Dances (i.e. for which Gets Burnt is the effect) because:

Intervention

A. You only observed women
B. You did not assign values to Dances
C. You left out other relevant variables
D. You did not assign values to Sober

19. Consider the causal system {Allergies[yes, no], Pollen in Air[yes, no], Spring [yes, no], Sneezes[yes, no]}, where Sneezes is the effect. Give an example of a test pair for Spring.

Test Pair-Multiple possible answers, one follows:

Allergies	Pollen	Spring	Sneezes
Yes	Yes	Yes	?
Yes	Yes	No	?

20. Consider the following causal system: {Sneezes[yes, no], Blinks[yes, no], Turned Away[yes, no], and Sees Shooting Star [yes, no]} in which Sees Shooting Star is the effect. Please draw a response structure diagram in which only Blinks and Sneezes are direct causes of Sees Shooting Star.
Direct Causation-Multiple possible answers, one follows:

Sneezes	Blinks	Turned Away	Sees Shooting Star
Yes	Yes	Yes	Yes
Yes	Yes	No	Yes
Yes	No	Yes	No
Yes	No	No	No
No	Yes	Yes	Yes
No	Yes	No	Yes
No	No	Yes	Yes
No	No	No	Yes

21. If you take two people from a population that exhibits Response Structure Uniformity, which ONE of the following is true?
Response Structure Uniformity

A. They will differ with regards to only one potential cause
B. If they have the same values for the potential causes, they will have the same value for the effect
C. They must have the same values for all potential causes.
D. They will show the same value for the effect.

22. Consider the set of variables {Just Ran One Mile [yes, no], Exhausted [yes, no]}, where Exhausted is treated as the effect. Which of the following populations is LEAST likely to have response structure uniformity in this case?
Choose exactly one of the following:
Response Structure Uniformity

A. A group of elite Kenyan marathon runners
B. A random sample taken from Irish and Kenyan populations
C. A mixture of elite Kenyan and elite Irish marathon runners
D. A group of obese people

23. What is a response structure? Give the definition with regards to test pairs, causal assignments, or the values of variables.
Response Structure

Quiz C

1. How is direct causation defined in the material you have studied in this experiment? Give the definition with regards to test pairs, causal assignments, or the values of variables.
Direct Causation

2. How is a causal assignment created?
Causal Assignment

3. If a population has Response Structure Uniformity, which ONE of the following is true?
Response Structure Uniformity

A. All members will have the same causal assignment
B. Members can only differ with regards to the value of one potential cause
C. Given the same causal assignment, all members will have the same value for the effect
D. Members will all show the same value for the effect

4. Consider this response structure:

Causal Assignment	Flu Shot	Average Sleep per Night	Avoided Flu
1	Yes	Yes	Yes
2	Yes	No	Yes
3	No	Yes	Yes
4	No	No	No

Which of the following is a test pair for Average Sleep per Night?
Choose exactly one of the following:
Test Pairs

A. 1 and 2
B. 1 and 3
C. 2 and 4
D. 2 and 3

5. Having the sickle cell gene in a population's gene pool can be beneficial to survival in a tropical climate, where it protects against malaria, but has no such beneficial consequences outside of tropical climates. This is most clearly an example of differences with regards to which of the following:
Causal Relativity-Background Conditions

A. Response Structure
B. The Causal System
C. Potential Causes
D. Background Conditions

6. Consider the following causal system: {Tired[yes, no], On Couch[yes, no], Watched Scary Movie[yes, no], Sleeps[Yes, No]}, where Sleeps is the effect. Give an example of a test pair for On Couch.
Test Pairs- Multiple possible answers, one follows:

Tired	On Couch	Watched Scary Movie	Sleeps
Yes	Yes	Yes	?
Yes	No	Yes	?

7. When do two potential causes interact?
Interaction

8. You invite two friends over for a dinner party at your house, where there are cats. One of them has allergies, and has an allergic reaction, while the other does not have allergies and is unaffected. This is due to a difference in which of the following:
Causal Assignments and Response Structures

A. Test Pairs
B. Causal Assignments
C. Interactions
D. Response Structures

9. According to the module, claims about direct causal relationships between variables are relative to what two major things?
Causal Relativity

       a._**Causal system (set of variables)**_____
       b._**Background conditions**_____

10. Consider the following causal system: {Sweats profusely[yes, no], showers[yes, no], smells[yes, no]} where smells is the effect. Please draw a response structure diagram in which sweats and showers interactively cause smells.
Interaction- *Multiple possible answers, one follows:*

Sweats	Showers	Smells
Yes	Yes	No
Yes	No	Yes
No	Yes	No
No	No	No

11. What is a response structure? Give the definition with regards to test pairs, causal assignments, or the values of variables.
Response Structure

12. Consider the causal system {Drinks Beer [yes, no], Grades [good, poor]}.
Suppose that you are interested in the influence of Drinks Beer on Grades. You observe your friends Herb and Sue, and you carefully record your observations in the following table:

	Drinks Beer	Grades
Herb	Yes	Poor
Sue	No	Good

This is not a response structure for Grades (i.e. for which Grades is the effect) because:
Choose exactly one of the following:
Intervention

A. Not enough variables have been included
B. Not everyone who drinks beer gets good grades
C. You have observed the value of Drinks Beer, not assigned it
D. You have observed the value of Grades, not assigned it

13. What is a test pair?
Test Pairs

14. With respect to which of the following causal systems is Presses button a direct cause of Changes Channel?
Choose exactly one of the following:
Causal Relativity-Set of Variables

A. Presses button[yes, no] TV receives signal[yes, no], Changes Channel[yes, no], Universal Remote[yes, no]
B. Presses button[yes, no], Changes Channel[yes, no], Universal Remote[yes, no]
C. TV receives signal[yes, no], Changes Channel[yes, no]
D. Presses button[yes, no] TV receives signal[yes, no], Changes Channel[yes, no],

15. What do we mean when we say a population has Response Structure Uniformity?
Response Structure Uniformity

16. Consider the Causal System {In Sun for 4 or more hours, Uses Suntan Lotion, and Gets Sunburned}. Imagine that you are interested in the effect of being out in the sun for more than 4 hours and using suntan lotion on getting sunburned. At a pool party, you tell your friend to put on suntan lotion and record how long he spends in the sun and whether he gets burnt. This is not a causal assignment because:

Choose exactly one of the following:
Intervention

A. You only observed how long he was out in the sun
B. You could not assign a value to "gets sunburned"
C. You required him to use suntan lotion
D. You only tested one person

17. If variable X interacts with variable Y to produce an effect Z, then which ONE of the following MUST be true?
Interaction

A. Y has the same value in all test pairs for X
B. There is at least one test pair for X across which the effect Z differs
C. Z has the same value in all test pairs for X
D. Z differs in all test pairs for X

18. Consider the following causal system: {exercised[yes, no], sleep deprived[yes, no], and tired[yes, no]}, where tired is the effect. Please draw a response structure diagram in which only exercised is a direct cause of tired.

Direct Causation- Multiple possible answers, one follows:

Exercised	Sleep Deprived	Tired
Yes	Yes	Yes
Yes	No	Yes
No	Yes	No
No	No	No

19. Consider a causal system involving 3 variables: {4 hours in full midday sun [yes, no], Used Sunblock [yes, no], Got Sunburn [yes, no]}. Which of the following populations will be MOST likely to exhibit Response Structure Uniformity for the effect variable: Got Sunburn?

Choose any number of the following:

Response Structure Uniformity

A. 5 dark skinned Kenyan men and 5 fair skinned Irish men.
B. A random sample of 100 people from New York City
C. 10 dark skinned Kenyan men.
D. A random sample of 10 Kenyan men

20. Consider the claim: "Drinking at least 3 glasses of tap water a day will improve your health." Which of the following are background conditions that this claim must be judged against (choose one):

Causal Relativity-Background Conditions

A. Whether your community has enough water
B. Whether your community's water supply contains toxic chemicals.
C. Whether your community is religious or not.
D. None of the above.

21. Given a causal system with 3 potential causes, <u>two</u> of which have 3 possible values (Good, Average, and Poor) while the other <u>one</u> has 2 possible values (Yes, No), how many causal assignments will be in a response structure for the causal system?

Causal Assignments-Number

18

22. Consider the following causal system: {Allergies[yes, no], Pollen in Air[yes, no], Spring [yes, no], Sneezes[yes, no]} where Sneezes is the effect. Please draw a response structure diagram in which only Allergies and Spring are direct causes of Sneezes.

Direct Causation- Multiple possible answers, one follows:

Allergies	Pollen	Spring	Sneezes
Yes	Yes	Yes	Yes
Yes	Yes	No	No
Yes	No	Yes	Yes
Yes	No	No	No
No	Yes	Yes	Yes
No	Yes	No	Yes
No	No	Yes	Yes
No	No	No	Yes

23. Consider the causal system involving: {Studies [yes, no], Drinks Beer [yes, no], Gets Enough Sleep [yes, no], and Grades [good, poor], and the following Response Structure for Grades:

Assignment	Studies	Drinks Beer	Gets average sleep	Grades
1	Yes	Yes	Yes	Poor
2	Yes	Yes	No	Poor
3	Yes	No	Yes	Poor
4	Yes	No	No	Good
5	No	Yes	Yes	Good
6	No	Yes	No	Good
7	No	No	Yes	Poor
8	No	No	No	Poor

Which of the following are test pairs for Drinks Beer?
Choose exactly one of the following:
Test Pairs

A. 1 and 2
B. 5 and 6
C. 1 and 3
D. 5 and 8

Appendix D: Basic Experimental Survey Items

1. Would you use the highlighting interface you used in this study if you taking an online course for credit?

      __ Yes   __ No

Comments?

_____

_____

_____

_____

2. How pleasant was this highlighting interface to use?

Very Frustrating	Moderately Frustrating	Mildly Frustrating	Neutral	Mildly Pleasant	Moderately Pleasant	Very Pleasant
1	2	3	4	5	6	7

3. On a scale of 1 to 7, how able were you to accomplish your highlighting goals using this interface?

Not At All						Completely
1	2	3	4	5	6	7

4. Do you believe the highlighting interface had any influence on how much you learned?

Strongly Reduced Learning	Moderately Reduced Learning	Mildly Reduced Learning	No Effect	Mildly Increased Learning	Moderately Increased Learning	Strongly Increased Learning
1	2	3	4	5	6	7

5. Describe your three favorite things about using the highlighting interface in this study.

     1. _____

     _____

     2. _____

     _____

     3. _____

     _____

6. Describe your three least favorite things about using the highlighting interface in this study.

     1. _____

     _____

     2. _____

     _____

     3. _____

     _____

7. We have noticed that, while reading text on the computer, some people select what they are reading with the mouse. If you do this, please let us know what this helps you accomplish.

     __I do not select text unless I plan to use it (e.g. copy paste, highlight)

     __It helps me pay attention

     __It helps me focus on specific parts of the text

     __I select text when the type is difficult to read

     __I select text as a bookmark to help me remember a location in the text

     Other (please describe):_____

     _____

     _____

Appendix E: Note-Taking/Highlighting Survey Items

1. What are your overall goals when taking notes on or highlighting text? Please rank them for both written notes and highlighting, with 1 being the most important. If you feel there

Goal	Notes	Highlighting
Pay Attention		
The process helps me learn		
To be able to review more quickly		
To ensure I review the most important material		
To use when interacting with other people		
Other		

_____

Comments?_____
_____
_____
_____

2. We are interested in understanding what type of material you take notes on or highlight when you are reading. Please **rank** all that apply, with 1 being the most important type of material.

Material	Notes	Highlighting
Definitions		
Key Words/Terms		
Key Ideas or phrases		
Good Examples		
Text that helped me understand the concepts		
Good summaries		
Things I don't understand		
Other _____		

_____

Appendix F: Design Study-Rating Interfaces

1. How pleasant was this highlighting interface to use?

Very Frustrating	Moderately Frustrating	Mildly Frustrating	Neutral	Mildly Pleasant	Moderately Pleasant	Very Pleasant
1	2	3	4	5	6	7

2. Briefly state the 3 things you LIKED most about the interface

1. _____
2. _____
3. _____

3. Briefly state the 3 things you DISLIKED most about this interface

1._____
2._____
3._____

4. Do you have any additional comments regarding the interface you used?

_____
_____
_____
_____

Appendix G: Design Study-Rating Dimensions

⇒ *Students were provided space to comment on individual questions*

We are interested in your responses to the different interfaces you used to highlight the material. Please answer the following questions to the best of your ability.

1. Please Rank the highlighting interfaces you just used, from favorite to least favorite:

Favorite   _____   _____
Middle   _____   _____
Least Favorite   _____   _____

2. Do you prefer to initiate the highlight action (e.g. click a button) before or after you make a selection?

Strong-Before	Moderate-Before	Mild-Before	Neutral	Mild-After	Moderate-After	Strong-After
1	2	3	4	5	6	7

3. Once you activate the highlight tool, would you prefer that only the next selection (*Only One*) you make be highlighted, or that every subsequent (*All Subsequent*) selection be highlighted until you click the tool again.

Strong-Only One	Moderate-Only One	Mild-Only One	Neutral	Mild-All Subsequent	Moderate-All Subsequent	Strong-All Subsequent
1	2	3	4	5	6	7

4. **Do you prefer that the toolbar (menu bar with highlight and undo buttons) be available at all times, or only when you have made a selection?**

Strong-Always	Moderate-Always	Mild-Always	Neutral	Mild-Selection	Moderate-Selection	Strong-Selection
1	2	3	4	5	6	7

5. **Do you prefer to use a visible toolbar to make a highlight, or another method, such as clicking on a selection you have made, pressing a key, or using the context menu?**

Strong-Visible	Moderate-Visible	Mild-Visible	Neutral	Mild-Other	Moderate-Other	Strong-Other
1	2	3	4	5	6	7

6. **If you are given a toolbar to highlight where do you prefer that it be placed? Circle one.**

   a. Top   b. Side

How strong is this preference?

Not at all strong	Mildly strong	Moderately strong	Very strong
1	2	3	4

7. **If you are given a toolbar that only pops up when you make a selection, where do you prefer that it be placed? (please rank)**

   ___ Top
   ___ Side
   ___ Where the mouse is located
   ___ Other(please describe)_____

How strong is this preference?

Not at all strong	Mildly strong	Moderately strong	Very strong
1	2	3	4

8. How do you prefer to access the tool that modifies (deletes) the highlight you have created? Please Rank.

___ Mouse over the highlight
___ Click on the highlight
___ Right click on the highlight
___ Other (please describe)_____

How strong is this preference?

Not at all strong	Mildly strong	Moderately strong	Very strong
1	2	3	4

9. Did you like that when you clicked on a sentence, the interface automatically selected that sentence?

Strong-Dislike	Moderate-Dislike	Mild-Dislike	Neutral	Mild-Like	Moderate-Like	Strong-Like
1	2	3	4	5	6	7

10. Do you have any additional comments regarding any of the interfaces you used today?

Appendix H: Highlighting Tools

Name	Interaction Description
YAWAS	After making a selection, the user accesses the context menu for the selection and clicks highlight.
Annotizer	After making a selection, the user clicks a highlight button in a toolbar on the left side of the browser
Microsoft Word	Users can highlight material after making a selection either by clicking a button in the toolbar at the top of the screen, or by using the context menu.
Conote	This application defines specific points in the material the user clicks on to create a highlight.
Annoty	This application offers a toolbar in the side of the screen. To create a highlight, the user must type the text they want to highlight into a textbox in the toolbar.
Office Web Discussions	Clicking a button at the bottom of the screen places a set of buttons in the content. Clicking on these buttons allows users to comment on them.
iMarkup	This interface provides several methods of recording notes. Users can first click a button in a toolbar at the top of the screen and then click a paragraph, which becomes highlighted. Users can also first select text, then use the context menu to highlight the selected text.
EDUCOSM	After making a selection, the user accesses the context menu for the selection and clicks highlight.
Gibeo.net	When a user selects text, a toolbar appears with a highlight button. Pressing the button highlights the text.
Third voice	After making a selection, the user clicks a highlight button in a toolbar on the top of the browser
Arakne environment	After making a selection, the user accesses the context menu for the selection and clicks highlight.
critlink	After making a selection, the user clicks a highlight button in a toolbar on the top of the browser
diane	To make a highlight, users first click a button and then select the text.
active annotations	Clicking on the annotatation button at the top makes every word in the document clickable. Clicking on the word highlights it.
Annotator	This interface requires students to open a separate browser to highlight the text. Intended for commenting.
Multivalent Documents	After making a selection, users click on the highlight entry in a list in the menu bar.
diigo	When a user selects text, a toolbar appears with a highlight button. Pressing the button highlights the text. Users can also use a button at the top of the screen after selecting text
marginalia	After making a selection, users can highlight it by either clicking a button on the side, clicking on the selection itself, or pressing a key.
JKN	When users click on a word, they are given a popup that allows them to comment on the word. Commenting highlights the word.
Annozilla	After making a selection, the user can highlight it by either pressing a button on a toolbar on the side, or by using an entry in the context menu

Shared Copy	Users "pickup" a highlighter by clicking a button on the top of the browser. Every selection the student makes is turned into a highlight until the highlighter is put down by clicking on the button again.
Hylighter	After making a selection, the user clicks a highlight button in a toolbar on the left side of the browser

Appendix I: Highlighting Design Study: Tools and Assignments

Name	Informal	Time	Visible	Duration	Contextual	Position
Pickup Highlighter	Pressing button "picks up" highlighter, so every subsequent selection becomes a highlight. Another button press releases the highlighter.	pre	yes	permanent	no	left
Press Button Before	To create a highlight, press button, then make selection.	pre	yes	single	no	top
Press Key Before	To create a highlight, press "h" key, then make selection	pre	no	single	no	-
Press Button After	To create a highlight, make a selection, then press button	post	yes	single	no	left
Press Key After	To create a highlight, make a selection, then press "h" key	post	no	single	no	-
Toolbar Follows Mouse	After making a selection, a highlight button appears near the selection. Press the button to turn selection into highlight.	post	yes	single	yes	mouse
Press Contextual Button	After making a selection, toolbar appears at top of screen. Press button to highlight selection	post	yes	single	yes	top
Context Menu	After making a selection, user can use the context menu to select "highlight".	post	no	single	yes	-
Click Selection	After making a selection, clicking on it highlights the selection	post	no	single	yes	-

Participant	Phase	Tool1	Tool2	Tool3	Contextual	Visibility	Location	Timing	Duration
1	1	Pickup Highlighter	Toolbar Follows Mouse	Click Selection	yes	yes	mouse-left	yes	yes
2	1	Press Key After	Context Menu	Press Button Before	no	yes	top-left-mouse	yes	no
3	1	Press Contextual Button	Press Button After	Press Key Before	yes	yes	top-left	yes	no
4	1	Press Key Before	Context Menu	Toolbar Follows Mouse	yes	yes	no	yes	no
5	1	Click Selection	Press Button Before	Press Button After	no	yes	mouse-top-left	yes	no
6	1	Pickup Highlighter	Press Key After	Press Contextual Button	yes	yes	top-left	yes	yes
7	2	Press Button After	Toolbar Follows Mouse	Pickup Highlighter	yes	yes	left-mouse	yes	yes
8	2	Toolbar Follows Mouse	Pickup Highlighter	Press Button After	yes	yes	left-mouse	yes	yes
9	2	Pickup Highlighter	Press Button After	Toolbar Follows Mouse	yes	yes	left-mouse	yes	yes
10	2	Press Button After	Toolbar Follows Mouse	Pickup Highlighter	yes	yes	left-mouse	yes	yes
11	2	Toolbar Follows Mouse	Pickup Highlighter	Press Button After	yes	yes	left-mouse	yes	yes

Appendix J: Data Tables

The following sections report data for each study described in this thesis.

Study 1: Handwriting and Text-Editing

The first study followed a counterbalanced within subjects design, with each subject handwriting notes while studying one module, and using a text-editor to take notes in the other module.

| Measure | HANDWRITING | | | | TEXT EDITING | | | |
| | Module 1 | | Module 2 | | Module 1 | | Module 2 | |
	Mean	StdErr	Mean	StdErr	Mean	StdErr	Mean	StdErr
Time (minutes)	74.03	7.60	38.63	4.94	69.09	7.60	44.08	4.94
Number of Notes	33.29	4.80	19.86	4.36	30.43	4.80	17.57	4.36
Words	421.29	127.27	221.86	115.96	564.71	127.27	485.14	115.96
Own Words	187.43	47.80	32.71	32.46	268.25	47.80	155.85	32.46
Verbatim Words	214.43	128.23	57.57	114.19	287.29	128.23	440.14	114.19
Quiz	77.38%	3.88%	75.32%	7.48%	73.81%	3.88%	70.13%	7.48%

Study 2: Copy-Pasting

The second study followed a between subjects design, with each subject using one tool to take notes on one module.

Measure	PAPER		TOOL- NO PASTE		TOOL- PASTE	
	Mean	StdErr	Mean	StdErr	Mean	StdErr
#Subjects	19	N/A	15	N/A	18	N/A
Time (minutes)	82.16	5.11	82.49	5.95	85.82	5.11
Words	418.74	88.84	354.60	99.99	814.56	91.28
Ideas	49.79	6.40	30.20	7.20	53.89	6.57
Key Ideas	9.21	0.48	7.73	0.54	7.94	0.49
Wordiness-Ideas (words per idea)	8.91	0.80	11.54	0.90	14.42	0.82
Ideas-Own Words	20.89	2.62	12.07	2.95	11.44	2.69
Ideas-Verbatim	16.32	5.00	13.33	5.62	40.17	5.13
Immediate Post-test, Multiple Choice	71.85%	3.45%	77.26%	3.75%	73.94%	3.75%
Delayed test, Multiple Choice	71.50%	4.74%	75.72%	5.15%	72.56%	5.15%
Review test, Multiple Choice	76.10%	4.04%	76.58%	4.39%	77.00%	4.39%
Immediate Post-test, Free Response	52.92%	4.99%	49.10%	5.43%	56.77%	5.43%
Delayed test, Free Response	52.98%	5.55%	57.78%	6.03%	46.36%	6.03%
Review test, Free Response	63.95%	5.28%	63.40%	5.74%	58.19%	5.74%

Study 3: Intervening On Selection

The third study followed a between subjects design, with each subject using one tool to take notes on one module.

Measure	TOOL-NO PASTE		TOOL- PASTE		RESTRICTED		SELECT	
	Mean	StdErr	Mean	StdErr	Mean	StdErr	Mean	StdErr
#Subjects	18	N/A	17	N/A	18	N/A	17	N/A
Time (minutes)	59.00	3.99	47.19	4.10	41.63	4.16	52.55	4.97
Words	590.75	109.00	1172.71	112.82	393.53	113.52	574.40	135.51
Ideas	44.85	6.72	72.83	6.96	26.23	7.00	36.94	8.36
Key Ideas	9.33	0.30	9.34	0.31	8.28	0.31	7.66	0.38
Wordiness-All Ideas (words per idea)	13.53	0.74	16.10	0.77	15.80	0.78	14.59	0.93
Wordiness-Key Ideas (words per idea)	18.20	0.82	23.27	0.85	19.68	0.85	16.84	1.02
Immediate Post-test	51.59%	4.80%	54.75%	4.90%	35.93%	4.80%	43.04%	4.90%
Delayed test	51.00%	5.00%	51.30%	5.20%	38.40%	5.00%	42.50%	5.10%
Review test	60.30%	5.40%	66.10%	5.50%	46.30%	5.30%	49.00%	5.50%

Study 4: Restricting Selection

The fourth study followed a between subjects design, with each subject using one tool to take notes on one module.

Measure	CLICK SELECT		NEW REC		NONE		UNRESTRICTED	
	Mean	StdErr	Mean	StdErr	Mean	StdErr	Mean	StdErr
#Subjects	13	N/A	13	N/A	12	N/A	12	N/A
Time (minutes)	44.69	5.384	43.69	5.384	39.23	5.384	44.16	5.6
Ideas	44.84	9.92	47.76	9.92	N/A	N/A	7.75	10.32
Key Ideas	8.77	0.357	9.15	0.357	N/A	N/A	8.75	0.372
Portion Alone-All Ideas	96.30%	3.90%	91.20%	3.90%	N/A	N/A	74.10%	4.10%
Immediate Post-test	70.62%	10.36%	89.87%	10.55%	111.63%	10.92%	89.80%	10.56%
Delayed test	82.37%	10.36%	88.91%	10.55%	107.32%	10.92%	81.85%	10.56%
Review test	98.44%	10.36%	97.68%	10.55%	116.30%	10.92%	114.02%	10.56%

Study 5: Highlighting vs. Copy-Paste

The fifth study followed a between subjects design, with each subject using one tool to take notes on one module.

Measure	HIGHLIGHTING		HIGHPAD		NONE		PASTE	
	Mean	StdErr	Mean	StdErr	Mean	StdErr	Mean	StdErr
#Subjects	15	N/A	12	N/A	12	N/A	12	N/A
Time (minutes)	45.21	3.37	34.73	4.25	49.08	3.66	50.96	4.92
Ideas	69.23	10.54	69.84	12.24	N/A	N/A	33.45	12.67
Key Ideas	8.98	0.32	9.31	0.37	N/A	N/A	8.69	0.38
Portion Alone-All Ideas	50.08%	6.36%	29.96%	7.39%	N/A	N/A	54.62%	7.64%
Portion Alone-Key Ideas	84.41%	2.80%	78.63%	3.25%	N/A	N/A	90.47%	3.37%
Immediate Post-test, Multiple Choice	50.43%	6.71%	52.70%	8.46%	50.57%	7.29%	58.64%	9.79%
Delayed test, Multiple Choice	52.58%	6.44%	51.16%	8.12%	56.53%	6.99%	64.71%	9.39%
Review test, Multiple Choice	54.31%	6.22%	62.20%	7.85%	59.19%	6.76%	70.22%	9.08%
Immediate Post-test, Free Response	39.66%	6.49%	52.48%	8.18%	45.51%	7.05%	49.08%	9.47%
Delayed test, Free Response	39.51%	5.16%	43.57%	6.51%	45.51%	5.61%	47.89%	7.53%
Review test, Free Response	48.51%	6.33%	58.12%	7.99%	38.46%	6.88%	65.57%	9.24%