The Annotators' Perspective on Co-authoring with Structured Annotations

by

Yamin Htun

B.Sc., Texas Christian University, 2005

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

The Faculty of Graduate Studies

(Computer Science)

The University of British Columbia

September 2007

Abstract

In asynchronous collaborative writing, annotations play an important role as a communication medium among co-authors. Research has shown that grouping related annotations together can help those who review an annotated document by reducing their workload and raising the accuracy of their reviewing. Less is known about the impact on users who create such structured annotations — the annotators. The research reported in this thesis had three goals: (1) to better understand current annotation creation practices, (2) to explore how structuring would be used by annotators, both the structuring process and the resulting types of structure, and (3) to evaluate the impact on annotators of having to create structured annotations. We conducted three studies to address each of these goals in turn. The first study was an observational study which strengthened our understanding of how annotators use existing tools to communicate document-related information in the form of annotations. That study revealed annotation practices that could benefit from additional structuring support, such as annotators describing how annotations in a document relate to each other. Our second study used a paper prototype system that supported annotation grouping to investigate how annotators would structure annotations, if given the option. Common behaviour that emerged was the grouping of thematically related annotations, as well as the grouping of annotations specifically targeted to a given co-author. The study also uncovered a range of temporal approaches to structuring annotations, such as top-down and bottom-up grouping. Based on the first two studies, we incorporated a light-weight implicit structuring approach based on tagging into our annotation model and then implemented an extended version of a high-fidelity prototype that supports structured annotation, including tagging. We used the prototype in our third study, a controlled experiment, which compared the impact of structured annotations relative to unstructured annotations. Participants in that study perceived structured annotations to be worth the additional workload required. The study further suggested that the bottom-up grouping approach complements the top-down approach in describing relationships amongst annotations in a document.

Contents

List of Tables

List of Figures

Acknowledgements

First, I would like to thank my supervisors, Dr. Joanna McGrenere and Dr. Kellogg Booth. This work would not have been possible without their guidance and support. I am grateful to Joanna for always providing detailed and thorough feedback. Her understanding and personal guidance has been of great value for me. I feel very fortunate and honored to have had Kelly as a co-supervisor. His wide knowledge and great ideas have had a remarkable influence on this work. I am also grateful to my second reader Dr. Steve Wolfman for his insightful feedback.

I would like to acknowledge my peers and friends for their feedback and support. In particular, I would like to thank Clarence and Zephyr, who were willing to lend an ear and offer advice. I would also like to express my gratitude to Myo, Hlaing and Mark, whose presence and friendship helped make my life outside of work truly enjoyable. Myo deserves special thanks for always being there for me, and for being especially understanding during the past few months.

Last, most certainly not least, my thanks go to my family for their constant encouragement and unconditional love. I am very grateful to my parents, who have always believed in me, and inspired all my achievements. It is to them that I dedicate this work.

The research reported in this thesis was supported by The University of British Columbia and by the Natural Sciences and Engineering Research Council of Canada under the Discovery Grants program and the Strategic Network Grants program as a component of NECTAR, the Network for Effective Collaboration Technologies through Advanced Research.

Chapter 1

Introduction

1.1 Research Motivation

Collaborative writing refers to the process of two or more people working together to create a document. In most collaborative writing tasks, co-authors typically annotate drafts and pass the annotated documents back and forth. Annotations play an important role as a central communication medium. Most word processing systems, however, only support simple annotations (basic edits and comments), forcing valuable communication among group members to take place outside the shared document, most often in the bodies of emails to which the document is attached and sent between co-authors. This results in communication being disconnected from the document, causing unnecessary overhead and inefficiencies [8]. For example, co-authors often need to provide explicit navigation statements such as "see page 2, paragraph 3," or they need to copy and paste some referenced text from the document into e-mail messages. This separation of artifacts means that valuable information can easily get disregarded or misplaced. These difficulties can increase dramatically with only a few reviewing cycles.

To address the shortcomings with current annotation tools, Zheng et al. developed an annotation model that unifies all document-related communication: single annotations are anchored at a specific place in the document, general comments are essentially anchored to the whole document, and structured annotations are a grouping of one or more single annotations or general comments [43]. Structured annotations may have hierarchical structure (groups within groups). The structuring is intended to communicate meta-information, i.e., act as meta-comments, about the group of annotations (e.g., summaries of edits). Zheng et al. evaluated the effects

of structured annotations on users reviewing an annotated document (i.e., the "recipients"), and found efficiency and accuracy benefits when compared to unstructured meta-comments written in e-mail messages. The effects of structured annotations on "annotators," those who create the annotations, had not yet been explored, and is the focus of this thesis.

1.2 Research Goals

Our target population is distributed groups collaborating asynchronously during the editing and reviewing stages of co-authoring, and creating a large volume of annotations to communicate document-related information. Before assessing the impact of working with structured annotations on annotators, it was fundamental and essential to our research efforts to first attain a better understanding of annotators' work practices with existing tools, including pen-on-paper markup. We wanted to examine how annotations are used, both the annotating process and the resulting types of annotations to communicate document-related issues.

As an initial investigation of structured annotations' usability, we next had to understand how annotators would go about structuring annotations if structure was provided to them as an option. We wanted to explore how annotators would use structure, for example, to thematically group related annotations, or perhaps to group annotations that were specifically targeted to a given co-author. Additionally, we wondered whether annotators would create complex hierarchical structures or non-hierarchical "flat" structures, and how much complexity the additional structures would add to annotating tasks.

In terms of creating structured annotations, Zheng et al.'s work assumed a relatively heavy-weight approach, where users would create explicit annotation groups called "bundles" [43]. Bundles make structure explicit and are especially helpful in creating annotation groups associated with preconceived structures and categories. This has been called top-down processing in the literature [18]. The recent explosion of interest in tagging systems [22], however, suggests that a more light-weight approach to information or-

ganization might be preferable or at least complementary. Tags allow groups to be formed more implicitly and are helpful in bottom-up processing where structure emerges rather than having to be pre-defined. We decided to see if tagging could be an appropriate way to structure annotations, allowing for more light-weight structuring than other methods.

Assuming that users were going to structure their annotations to communicate meta-information about annotations, we wondered how this would impact the overall amount of information communicated among co-authors. We wanted to investigate whether users who had a tool for structuring annotations would create more meta-comments, our assumption being that creating structured annotations would be an easier way to provide meta-comments than doing so separately in the body of an e-mail.

In addition to the amount of communication, the impact of structured annotations on annotators' workload had to be explored. It was not known whether the overall workload for structuring annotations would be similar to users providing unstructured annotations and then having to compose a detailed e-mail with the equivalent meta-information. We also had to investigate whether annotators would perceive structured annotations to be worth the effort. It was crucial to investigate the costs (effort) and benefits tradeoff associated with structured annotations for annotators because, as Grudin noted, tools that have high costs will not be used unless those perceived costs are balanced by high perceived benefits to the people doing additional work [14].

1.3 Research Approach

In order to address the above research goals, we conducted our research in three phases. First, we conducted an observational study to better understand annotators' workflow with existing annotation tools, none of which provide any explicit support for structuring annotations. In particular, we observed the use of traditional pen-on-paper markup and the use of Microsoft Word, a popular commercial word processor. In the second phase, we sought to identify different structuring approaches, and the types of

structures that annotators would use. We explored these issues independent of the tools to create structured relationships (and their potential usability issues) by conducting an exploratory study with a paper prototype. Based on the first two studies, we then extended an earlier prototype designed to structure annotations [43] by incorporating bottom-up structuring (i.e., tagging). Lastly, we conducted a formal experiment in which the prototype system was compared to an equivalent system that did not support structure, and investigated the impact of structured annotations on workload and the amount of information communicated.

The Observational Study and the Experiment were conducted with writing tutors at The University of British Columbia who professionally annotate documents to help students with their writing skills. Having reviewed and annotated numerous documents of various lengths and types, they have experience communicating document-related information. We used experienced annotators because we wanted to understand how structuring annotations would be used to address different types of errors (e.g., syntax, semantics); less experienced writers tend to focus only on syntactic errors when reviewing documents [24]. For the Paper Prototype Study, graduate students were used for recruiting efficiency. In that study, annotations did not need to be created, only organized; the level of experience required was relatively lower than in our other two studies.

1.4 Research Contributions

This thesis documents work done to examine the effects of supporting structure on users who create annotations. From the Observational Study, we have strengthened our understanding of how annotators use existing tools (both digital and pen-on-paper markup) to communicate document-related information in the form of annotations. We have also identified common uses of structured annotations and different approaches to structuring annotations through the Paper Prototype Study. Our contributions include a light-weight implicit structuring approach based on tagging, and an extended version of an interactive prototype that supports structuring anno-

tations. With the prototype, we evaluated the use of structured annotation support for annotating documents and communicating document-related issues. From the Experiment, we showed that structured annotations are perceived to be worth the effort despite the additional workload, and that bottom-up and top-down approaches to structuring annotations are complementary so both should be supported.

1.5 Overview

This thesis comprises descriptions of three studies that were designed to address the research goals described earlier in this chapter. Previous work relevant to this research is summarized in Chapter 2. In Chapter 3, we describe the Observational Study, which was designed to better understand annotation creation practices. Chapter 4 describes the Paper Prototype Study, which was our initial investigation of how users might go about structuring annotations. Chapter 5 summarizes the key findings and implications that we derived from the first two studies, and discusses motivations for structuring and tagging. It then describes an extended version of a high fidelity prototype that supports structured annotations including tagging. Chapter 6 presents the Experiment, which examined the impact of structured annotations on workload and the amount of information communicated. Finally, Chapter 7 summarizes the main results in the thesis and discusses several areas for future research.

Substantial portions of this thesis appear in a conference paper submission jointly authored with my supervisors, Dr. Joanna McGrenere and Dr. Kellogg Booth.

Chapter 2

Related Work

In this chapter, we first describe general background on collaborative writing research. We then provide a more focused discussion of research on document-related communication in the form of annotations among co-authors, and we survey systems that provide support for structuring annotations. Our project is based on organizing annotations; hence, we also highlight key approaches to information organization, with a focus on tagging.

2.1 Collaborative Writing

Collaborative writing refers to the process of two or more people working together to create a document. Collaborative writing is a common practice; research has shown that more accurate and better quality documents can be achieved with the collective knowledge and expertise from co-authors [10, 26]. As part of collaborative work, collaborative writing can be summarized in four modes based on the degree of proximity, i.e., whether members work from the same location, and the degree of synchronicity, i.e., whether members work on the document at the same time. Table 2.1 illustrates the four modes of collaborative writing.

Researchers have investigated the overall co-authoring process [29, 31], and identified different activities involved in co-authoring: brainstorming, researching, planning, writing, editing and reviewing [29]. Many collaborative writing tools have been developed to support these different activities. Classic collaborative writing systems appearing in the research literature include SASSE [6], PREP [25], and Quilt [12]. More details about these and other co-authoring systems can be found in [29]. With increasing accessibil-

		Synchronicity	
		Same Time	*Different Time*
Proximity	*Same Location*	Face to Face	Asynchronous Same-Place
	Different Location	Synchronous Distributed	Asynchronous Distributed

Table 2.1: Collaborative writing work modes (adapted from Ellis et al. [11])

ity and pervasiveness of the Internet, commercial web-based collaborative authoring tools have also been developed such as: Collaboratus [21], WriteBoard [4], and Google Docs [3].

Despite significant development work over the past decade, collaborative writing tools remain underused, according to studies conducted by Kim and Eklundh [17], and Noel and Robert [26]. Their studies found that most people write asynchronously within small groups, and members use personal word processors, instead of specialized co-authoring tools to write collaborative documents. Group members communicate with one another using native annotation tool support within word processors, and external communication channels such as e-mail. However, communicating document-related issues separately in e-mail messages can cause inefficiencies as discussed in Chapter 1. Therefore, the goal of the work reported in this thesis was to provide comprehensive annotation support that would allow co-authors to communicate document-related information within a collaborative document.

In our review of the literature, we focus primarily on document-related communication in the form of annotations among co-authors.

2.2 Annotation Support in Co-authoring Contexts

Wojahn et al. studied the effects of annotation interfaces [41] on communication among co-authors. Not surprisingly, they found that difficulty in producing annotations often results in brief annotations with less elaboration.

Churchill et al. [8] developed a light-weight communication tool called "Anchored Conversations." The tool supports real-time communication in the context of collaborative documents by allowing conversation scripts to be anchored into specific parts of a document. This merges shared discussions and document artifacts, facilitating the establishment of context information for document-related conversations. However, we suspect that verbosity of full conversations may overload authors when retrieving the information.

Current commercial systems (such as Adobe Acrobat and Microsoft Word; see [42] for details), and the collaborative writing systems described earlier (such as SASSE and PREP) provide simple annotation features only. To enhance annotation support, richer annotation models have been developed [37, 43]. An activity-oriented annotation model [37] was developed and implemented in a web-based collaborative writing tool (called "PCAT") designed specifically for co-authoring and reviewing clinical trial protocols [38, 39]. In PCAT, annotations can have properties such as response deadline and urgency. Users can also assign each annotation to one of the model's pre-defined categories such as "question" or "reply." The system also supports general comments that are attached to the whole document and threaded annotations for iterative discussions among co-authors. Although the model extends basic annotation features, we suspect that pre-defined categories may be too rigid and limited to capture many of the activities involved in co-authoring.

The annotation model developed by Zheng et al., which motivated our research, unifies all document-related communication by supporting structured annotations [43]. The annotation model is implemented in a research prototype called the "Bundle Editor." Similar to PCAT, the Bundle Editor

supports threaded annotations and general comments. The system allows users to group annotations, and communicate meta-information in the form of annotation groups such as a list of to-do items, or a summary of edits. Groups in the "Bundle Editor" are not pre-defined; users can create any kind of annotation group, and they can add a name and an optional note to indicate the nature of the grouping. Although this model is the first to support annotation grouping in co-authoring contexts, the concept of annotation grouping is not entirely new; annotation systems with grouping support have been developed in other contexts. We describe some of these systems in the following section.

Structured Annotation Support in Other Contexts

The Knowledge Weasel system [20] features collaborative annotation with the purpose of capturing structural knowledge — the knowledge of relationships among a set of documents. The annotations in the system serve as links between the files (one as annotated source, and the other as annotating target). The system allows users to navigate annotations locally by following links, or globally by querying the annotations' attributes. It also allows collaborators to group related annotations to form a hyperlink network of related information resources.

Ovsiannikov developed a system called Annotator, an annotation tool for taking notes on published web pages or HTML documents [27]. The system allows users to store annotations in the form of atoms, clumps and notes. An atom is similar to a single annotation in our structured annotation model — selected text with a comment attached to it. A note is similar to a general comment and attached to the whole document. A clump is defined to be a set of semantically related atoms linked to the same annotation. Users can retrieve annotations from different documents and generate new clumps with query results. Hence, in Annotator, the grouping is largely across HTML documents.

In software development contexts, a tool named TagSEA was developed to support collaborative annotations among group members [35]. The tool

provides a light-weight mechanism to organize annotations by associating them with the same "tags" or keywords. For example, users can add keywords such as "bugs" or "performance" to annotations anchored at different sections of program files to group the annotations together. Users can then browse annotations through hierarchies of tags or filtering based on particular tags. The goal of this tool is to enhance coordination among members and capture important knowledge about source code.

To our knowledge, no formal evaluations have been reported for any of the above systems. Hence, little is known about the impact on users of having to create structured annotations. Our research is the first to assess the impact of supporting structure on users who create annotations in a collaborative writing context.

2.3 Information Organization Approaches

Our research also fits within the broader research area of information organization because we focus on organizing annotations. We provide a brief survey of the literature on information organization and discuss different organization approaches including tagging, which we identified earlier as a potentially more light-weight approach for structuring annotations.

Researchers have studied how people organize and manage information for future retrieval in different contexts such as e-mail [40], web bookmarks [5], and files [16]. Hierarchical or tree-like structural systems are the most common paradigm for filing and categorizing information [30]. However, hierarchical systems present several challenges in both the categorization and retrieval of information. Although multiple categorizations are supported in existing operating systems (with "shortcuts," "aliases," and "symbolic links"), they require extra effort and are not pervasive throughout systems. Hence, multiple categorization features are not commonly used [30]. The act of categorizing into a single group is cognitively difficult [30, 34]. Moreover, the path order dependence enforced by hierarchical systems restricts the way in which the information can be retrieved and requires users to remember the exact path [7].

In order to enhance the retrieval process, search engines have been developed to find desired information by typing in just a few keywords (e.g., Google Desktop, Apple's Spotlight). Despite the possibility of such search engines returning perfect results with a small set of keywords, the search tools require users to specify the keywords inherent in the desired information content, which is difficult, especially in collaborative contexts where information is created by other individuals. A diary study conducted by Teevan et al. [36] reported that keyword-based search engines were not commonly used when searching for information; a browsing strategy was used instead because it offered better understanding of the desired information and a sense of location during searches. Jones et al. [16] found in their study that categories helped users see the relationships within their information and to have a sense of control over the grouping and location of information.

Recently, tagging has grown popular as an alternative to hierarchical structuring and keyword searching for organizing information resources, especially in collaborative contexts. For example, del.icio.us [1] is a web-based tool for organizing and sharing bookmarks based on tagging, and flickr [2] is a similar tool for digital images. In such tagging systems, users assign meta-data or keywords to information resources. Traditionally meta-data is created by professionals (catalogers or authors) [22], but tagging systems allow ordinary users to describe and organize content with any vocabulary they choose. Tagging systems offer two major capabilities: (1) they allow users to add tags to information in a light-weight manner, and (2) tags serve as navigational aids for users to find and organize the information later [28]. The prevalence of a given tag in a system is visible to users through the display of its frequency. Users can easily access the annotations labeled with the same tag through a single click, thereby serving as a navigational aid.

Tagging systems emphasize user-defined keywords as "a fundamental organizational construct" [22]. Unlike hierarchical or tree-like structural systems, tagging systems do not require users to develop and agree on a hierarchy of structure to organize and retrieve information; instead they just need to have a shared understanding of a tag's meaning to achieve cooperation and shared value. Collaboration is encouraged by browsing and

searching shared tags that are relevant to one's interests. Moreover, navigation of information resources through shared tags is similar to conducting keyword-based searches, except that users are not restricted to the exact keyword terms inherent in information resources [13]. Tagging is claimed to require less cognitive workload from users than other information organization schemes [34]. Hence, information organization using a tagging approach could offer higher benefits at lower costs, and might impose fewer barriers for collaboration than other approaches.

Chapter 3

Observational Study to Understand the Annotating Process

3.1 Study Goals

In this chapter, we describe a small qualitative observational study conducted to help us understand annotators' workflow with existing annotation tools: Microsoft Word (a popular commerical word processor), and traditional pen-on-paper markup. Zheng et al., before developing their structured annotation model, conducted a study in which they collected collaborative artifacts retrospectively from users (co-authors of academic papers), and analyzed meta-comments in the e-mails [43]. They did not analyze annotations embedded in the documents, nor investigate the workflow involved in creating annotations. Little evidence existed about: (1) what process users use to annotate the document, and (2) what kinds of annotations are created in the document. We sought to understand these issues by observing users creating annotations in a document, and by analyzing the annotations created. We first describe the methodology of our Observational Study and then discuss the behaviors and practices observed.

3.2 Methodology

We conducted the study with writing tutors from The University of British Columbia. Recruited through online mailing lists, a total of 5 tutors (4

females) participated in the study. All participants had been tutoring academic writing skills for more than three years. Two participants had previously reviewed more than 200 documents, and three between 50 to 100 documents.

The study was designed for a single two-hour session. At the beginning of the study, participants were asked to review and annotate a four-page essay-style document (see Appendix A.2) with approximately 1,500 words using their method of choice; two participants annotated with a pen on the printed document, and three used Microsoft Word with its track changes and commenting functions. Participants were instructed to provide their feedback on the document as they would normally do in their tutoring practice. A simulated e-mail message window was also provided to all participants, allowing for additional document-related communication directed to the hypothetical recipient. Reviewing was followed by a semi-structured interview to probe the annotating practices observed. Appendix A.3 shows the interview questions. Two hours were required for each participant to complete the study. Participants were paid $40 for their participation.

3.3 Results

We report our findings from the study. We observed and took notes on all behaviors related to document reviewing and annotation creation. This allowed us to understand the process of creating annotations. We also collected the annotated documents at the end of the task and examined the annotations (example annotations are shown in Tables 3.1 and 3.2 at the end of this chapter). In addition, qualitative feedback provided during the interview was transcribed.

3.3.1 Multiple Passes

All participants made at least two passes through the document while reviewing. They made annotations about syntax issues (e.g., grammar) on their first pass. Then, they took another full pass or quick skim to check

semantic issues (e.g., argument structure) and to achieve an overview of document status, errors made, and remaining work to be done. They then wrote comments on those issues at the end of the document or in an e-mail message. For example, general comments such as *"Your writing is clear and easy to follow ...* ," and *"I found the piece well-written, and well-structured. Most changes are minor ...,"* were written at the end of the printed document or in the accompanying e-mail messages.

One participant wrote a summary of her review on each section of the document in an e-mail message. She included section titles of the document such as: "Introduction," "Body," and "Conclusion," and wrote overall feedback on each section such as *"Introduction goes from general to specific, good. But you wait too long to introduce groupware."* Summaries of repeated errors that occurred in the document were also included, such as *"GRAMMATICAL — some tendency to use unnecessary words, and word combinations ... Examples, 'decisions have to be made,' instead of 'decisions are made.'"* However, when asked during the interview, all participants mentioned that they did not want to spend a lot of time summarizing their review, which essentially involved repeating and referring to what they had already noted in the document.

3.3.2 Justifications

We found that all participants not only made suggestions for changes, but also occasionally provided an explanatory comment along with their annotations, particularly when a problem was encountered for the first time. For example, as shown in Table 3.1(i), an edit annotation suggesting a verb tense correction was accompanied by a comment explaining *"Stay in the same tense as the rest of the sentence."* All participants revealed that explanatory comments were added to help annotation recipients better understand the errors and the changes made to the document. When the same error was repeated in the document, participants made the edit changes to correct each instance of the error, but did not add an explanatory comment again. They expected recipients to refer back to the annotation at a previ-

ous occurrence of the problem. For example, one participant mentioned in the interview that, *"If there is no comment accompanying a [edit] change, it is because I have previously changed the same error and commented on it before."* We note that this is one place where structure may be beneficial: all instances of the same problem can be linked together to reduce ambiguity.

3.3.3 Local versus Global Comments

All participants gave two major types of feedback on the documents: "local" or sentence-level feedback, and "global" or document-level feedback. One of the participants explained:

> *"There are two different types of comments [that I usually make] — comments that are specific for a specific sentence — why this sentence is not working, or what have you done wrong, or things like that [that] need a little bit explanation. But generally comments are about larger issues — say for example, ... whether or not the language is fitting to the general audience. So, it is more of a general comment, so maybe I want to highlight a few ideas that relate to that comment."*

All participants created single annotations (annotations embedded in the document with a single anchor to specific content) to address sentence level or local issues. Regarding global issues, participants typically inserted a comment such as *"Example of non-academic language — read through for this sort of language"* as a single annotation to the place in the document where the problem was first realized. We note that with no explicit additional pointers, recipients would not necessarily be able to see all instances of a problem. This is another place where additional structure may be beneficial.

3.3.4 Tagging-like Behaviors

Two participants used a keyword association technique to efficiently point out errors in the documents. For example, one participant defined a keyword

coding as *"WC=Word Choice"* and added the keyword "WC" to every place where she found a wording problem, instead of writing more verbose comments repeatedly. We note that this is an instance of tagging-like behaviors where the tags are symbols or abbreviated notes.

3.3.5 Prioritization

Two participants occasionally highlighted the text anchors of some annotations using, for example, a yellow highlighter or italicized fonts. One participant explained that she did so to indicate the higher degree of urgency and importance of those annotations. The participants thus used highlighting to ensure that recipients would pay specific attention to the most important annotations that might otherwise be "buried" among other annotations in the document, a common problem with heavily annotated documents.

3.3.6 Reviewing Methods

Participants explained their method of choice (Microsoft Word vs. pen-on-paper) as follows. Three participants said that they preferred to annotate documents electronically because they could conveniently send or receive electronic copies of documents via e-mail. The two other participants mostly annotated printed copies of documents because they did not have access to electronic copies at the time of tutoring or because it was easier for them to draw and visualize some ideas or outlines of the documents on paper.

As the example in Table 3.1 shows, all three participants who used Microsoft Word recorded the edit changes they made to the document by using the "Track Changes" (TRK) feature, which automatically creates edit annotations. When adding comments to the document, we observed that these participants used two different methods: (1) using the commenting function of Microsoft Word, and (2) typing directly into the document text using special fonts (such as italics). One participant exclusively used the first method to add comments while another participant exclusively used the second method (he explained during the interview that he was not aware of the commenting feature in Microsoft Word). The remaining third partici-

pant used both methods; she felt that it was sometimes easier for her to type directly into the text than to click the "Insert Comment" button and add a comment. Annotations (both edits and comments) in Microsoft Word are displayed with a markup balloon in the right margin of the document near the anchored text. Participants described their concern with the display of annotation balloons in Microsoft Word; they were concerned that visual clutter resulting from displaying all annotations in the document might have recipients overwhelmed and intimidated with the amount of annotations in the documents.

The two participants who annotated printed documents marked on the document text to indicate the suggested changes. These participants also wrote some comments in between the lines or in the margins with their anchored text circled, underlined or enclosed by parentheses. The examples in Table 3.2 show samples of annotations made in the printed documents by the participants.

3.3.7 Other Communication Methods

The three participants who used Microsoft Word in our study mentioned that they had tried other methods such as online chatting and phone calls. Although these methods allowed for more interactive discussions and conversations, our participants did not like them because (1) those communications happened outside the document, requiring additional explicit navigational statements such as "see page 3, paragraph 2" to build a context for conversations, and (2) such interactive communication did not allow the participants to control their reviewing time sufficiently. For example, one participant who used MSN messenger shared her frustration with the online messaging method,

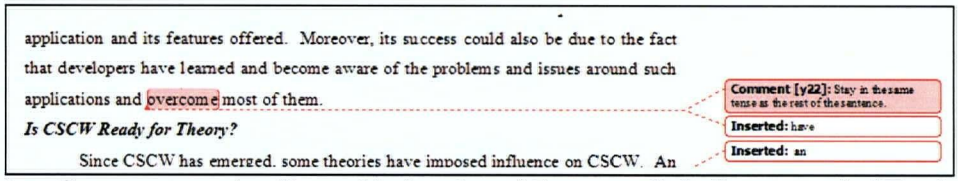
> *"Once I gave them my MSN [id], they keep talking to me day after day and everyday ... and asked me questions about their draft. It gets really frustrating. And [I have] to be the one to say 'OK, I can't help you anymore. I've got my own work to do.'"*

Hence, the participants chose not to use these methods anymore, and mainly used the annotation features of Microsoft Word and e-mail, in which they had more control over the length of time they allocated to reviewing tasks.

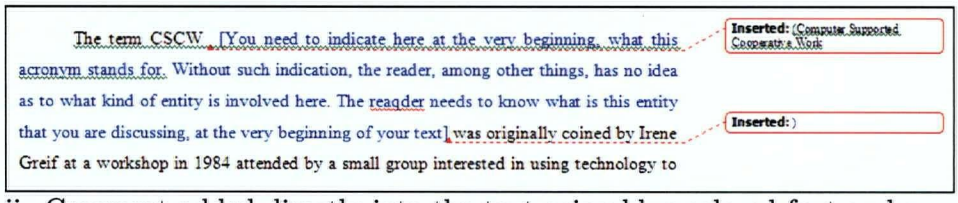
3.4 Summary

The Observational Study provided us with a better understanding of different types of annotations and workflow involved in their creation. While reviewing documents, participants gave feedback not only to improve the quality of the reviewed document, but also to enhance the writing skills of recipients by including explanatory comments along with some of their suggested edits. The findings also confirmed that gaps exist between current methods of annotating (specifically pen-on-paper markup and Microsoft Word) and annotators' needs: (1) annotators lacked an efficient means for describing relationships between annotations, and (2) they also lacked a keyword annotation feature, such as tagging, to allow for efficient feedback. We also learned that participants would like to have a priority scheme for highlighting important annotations, and an ink annotation feature for drawing and visualizing ideas in the document. These are features we had not investigated in our study but should be considered in future studies. Another interesting finding was that participants preferred to have more control over their reviewing time, and to keep the interactivity with recipients low.

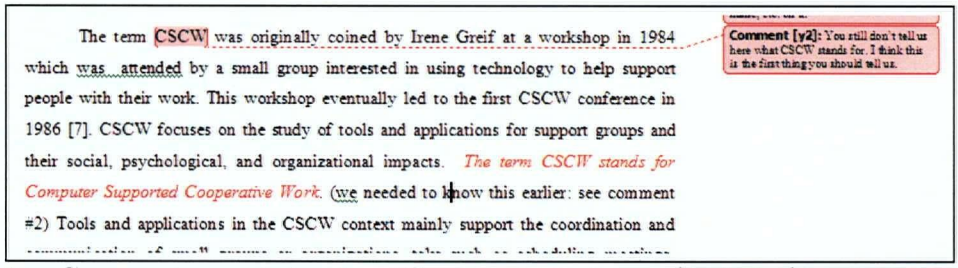
Having identified different places where structure may be beneficial, our next step was to examine how annotators might actually go about structuring annotations if structuring was provided to them as an option. In the next chapter, we describe a qualitative study involving a paper prototype in which we explore this issue.



i. Comment and edits added using "Comment" balloons and "Track Changes."



ii. Comment added directly into the text using blue colored font and surrounded by parentheses.



iii. Comments created using "Comment" balloons (first line) and typing directly into the text (see the second last line) with parentheses used to distinguish the comment from the document text.

Table 3.1: Examples of annotations created by participants using Microsoft Word.

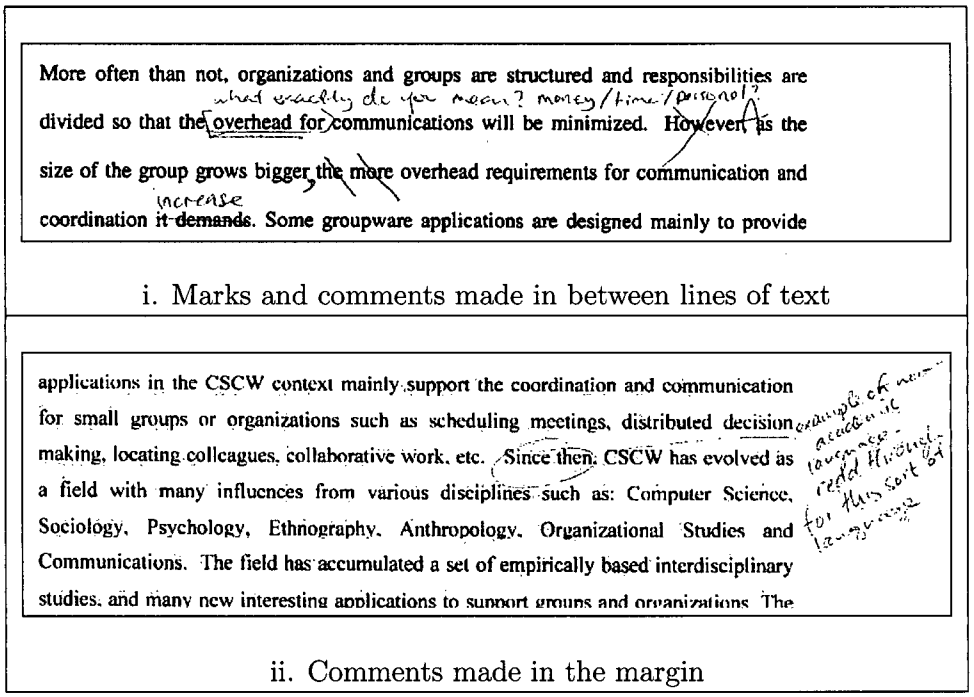


Table 3.2: Examples of annotations created by participants using a pen on a document printed on paper.

Chapter 4

Paper Prototype Study to Understand Annotation Structuring

4.1 Study Goals

Having confirmed annotators' needs for more structured annotations, we sought to understand what process and what types of structure annotators would use when provided with structuring as an option. More specifically, we wanted to assess: (1) the semantics of the structures created, (2) the approaches taken to create structure (top-down, bottom-up, or otherwise), and (3) the complexity of the structuring created in terms of the size of annotation groups and whether hierarchies (e.g., groups within groups) might be adopted. In order to mitigate the impact of any particular tool (and its potential usability issues), we elected to do a qualitative exploratory study with a paper prototype where grouping annotations amounted to essentially making little piles of paper annotations.

4.2 Methodology

The study was qualitative and exploratory in nature. We did not have specific hypotheses at the outset of the study.

4.2.1 Participants

A total of 8 people (5 females) participated in the study. Recruited through online mailing lists and newsgroups, they were all graduate students at The University of British Columbia: one from Zoology, three from Psychology, and four from Computer Science. A screening process ensured that all participants had co-authoring experience; five participants had co-authored more than ten documents, two between five and ten documents, and one less than five documents. Collectively, the participants had co-authored a wide range of documents including brochures, project reports, technical documentation, journal papers, and an encyclopedia chapter.

Five participants used a word processor (mainly Microsoft Word) everyday, two every 2–3 days, and one did so once a month. They all felt very confident about using their word processor, and only one participant had never used any annotation functions. Participants were paid $15 for their participation.

4.2.2 Task

Participants were asked to perform a task of organizing annotations in a document. They were instructed to assume the role of a co-author collaboratively writing a given document with two other co-authors who had expertise in different areas. The participants' task was to organize *pre-existing* annotations in the document, ones they had hypothetically just created, so that their co-authors could review the document efficiently and accurately.

The document consisted of 932 words and 42 annotations. Because we were interested in variability among users' grouping approaches and annotation groups, we provided the same document and annotations to all participants, who were asked not to add any new annotations. The scenario and annotations were designed with an outlook that different kinds of annotation groups could be created (e.g., based on types {edits, comments}, themes {tone, clarity}, or targeted co-author). The document was about understanding the effects of different types of music and volume on students' ability to study. For the most part, the content was general enough

for all participants to understand. It did contain descriptions of experiment methodology and analysis, which might be more accessible to Psychology students, something we only realized in retrospect. Appendices B.4 and B.5 show the instructions given to participants and the annotated document used in the study.

Participants were given a printed copy of the document with annotations displayed on the right margin and their anchored text highlighted. Separate identical copies of each annotation printed individually on small paper strips were made available so that participants could pile the strips together (and optionally paper-clip them) to make annotation groups. To identify a pile, a post-it sheet was placed on top for writing the annotation group's name and an optional note. Multiple copies of each annotation were made available so that participants could place an annotation into more than one group. Each annotation group was allowed to be nested under other groups in any hierarchical structure. Figure 4.1 shows a participant performing the annotation-organization task, and Figure 4.2 illustrates sample piles of annotation groups that were created.

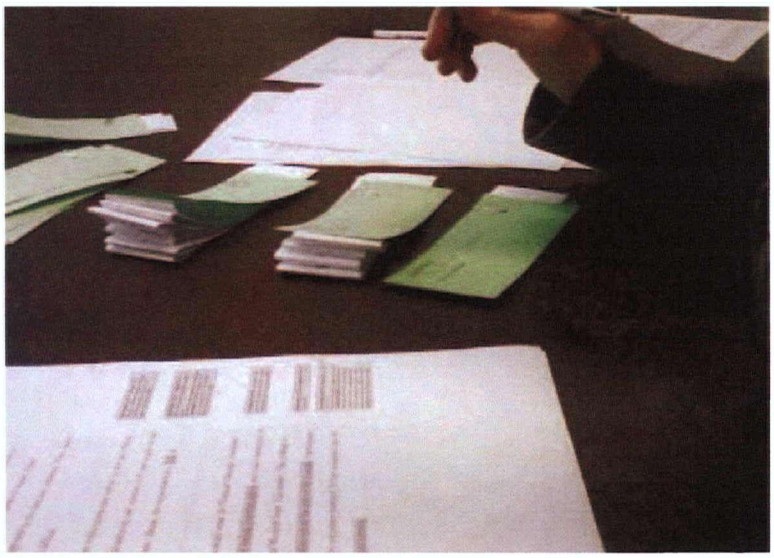


Figure 4.1: A participant performing the task with the paper prototype.

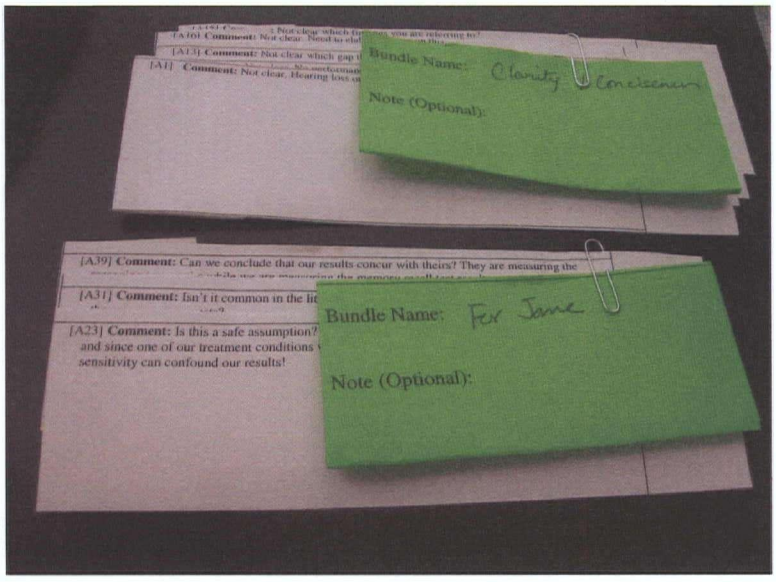


Figure 4.2: Sample piles of annotation groups created with green post-it notes identifying the name of the group (or bundle) and an optional note.

4.2.3 Procedure

The study was designed for a single one-and-a-half-hour session. A demographic questionnaire was followed by an information session on general concepts such as collaborative writing and then a training session on how to use the paper materials in the task. We were concerned that documents and annotations used in the training session might interact with the task performance; hence a list of shopping items needing to be organized was used in illustrating the paper materials. Appendix B.3 shows a detailed description of the training task. Participants were next asked to read an annotation-free version of the task document, after which they were given the annotated version. They had to perform the annotation organization task. For the first pass over the annotated document, participants were required to read the annotations in the order that they appeared in the document so as to simulate that they had themselves annotated the document in sequential order. They were, however, allowed to start grouping annotations at any point

during the task. A short questionnaire and a debriefing session were used to gain better insight into the grouping behaviors and preferences observed.

4.3 Results

We analyzed the recorded video of participants performing the task, and coded all behaviours related to the simulated reviewing (i.e., participants reading pre-existing annotations) and annotation group creation. This allowed us to understand the temporal patterns of annotation grouping. We also collected the "piles" of annotation groups at the end of the task to analyze their structure. In addition, qualitative feedback provided during the debriefing session was transcribed.

4.3.1 Temporal Approaches for Creating Groups

Participants followed different temporal sequences for grouping annotations: four dominant patterns emerged (as shown in Table 4.1) and are distinguished by the number of passes over the document that were made and when, with respect to those passes, annotation groups were created.

The participant who used the *pre-review* approach mentioned that the co-authoring scenario informed him of the annotation groups he wanted to create. This may have been an artifact of our study design, or may simply represent a difference in style, as none of our other participants followed this approach exclusively. The three participants who used the *post-review* approach said that seeing all annotations in the document before grouping helped them make their groups more consistent and manageable. They mentioned that they took *mental notes* of annotations of interest so that they could relocate them for grouping later. Two of these participants externalized their mental notes by adding keywords or notes to annotations on the printed document during their first read through, and then grouped annotations based on those keywords. The two participants who used the *during-review* approach stated that they grouped annotations "naturally" as occurred to them without explicitly having to think about grouping. Lastly,

	Description
Pre-review Num of pass: 1 (1 participant)	Participants formulated annotation groups prior to seeing annotations. Annotations that fit these pre-defined groups were later selected and associated with corresponding groups.
Post-review Num of pass: 2 (3 participants)	Participants read *all* annotations prior to formulating any groups. Once groups were defined, relevant annotations were associated with the groups.
During-review Num of pass: 1 (2 participants)	Participants organized annotations into groups while reading annotations, created new groups when existing ones were not appropriate for a given annotation.
Hybrid Num of pass: 2 (2 participants)	Different groups of annotations were created using different approaches stated above *(pre-, post-* or *during-* reviews)

Table 4.1: Temporal patterns of structuring annotations. (N=8)

the two participants who used the *hybrid* approach mentioned that they created "obvious" groups (such as typos and grammar) by using the pre-review approach and other groups by using other approaches. After grouping all annotations, four out of the eight participants (2: *hybrid*, 1: *post-review*, 1: *during-review*) also reshuffled some of their groups by merging or splitting.

4.3.2 Semantics of Annotation Groups

Annotation groups resulted from organizing disparate annotations throughout the document that were conceptually related. Most groups (82%) were *problem-based*: annotations were similar in the nature of the problems that they addressed (e.g., the group named "Tone" had annotations that highlighted and discussed inconsistent tone throughout the document). The three most common types of problem-based groups created by participants were clarity, grammar, and structural or organizational problems. The remaining groups (18%) were *recipient-based*: annotations that were to be

reviewed by an intended particular co-author (e.g., the group named "For Jane" with annotations that solicited Jane's expert knowledge).

4.3.3 Structures of Groups

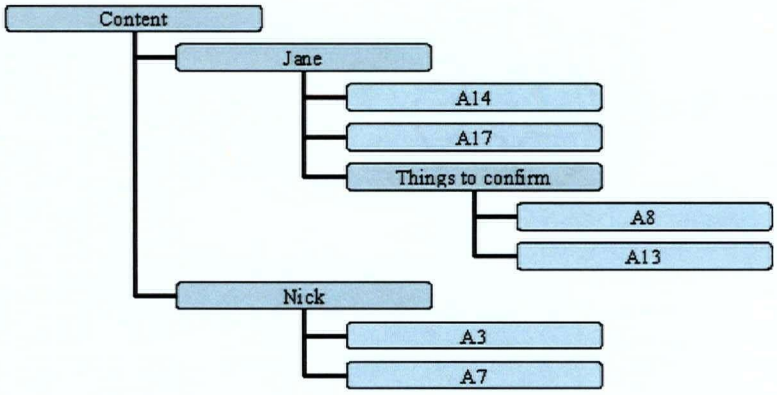


Figure 4.3: Sample annotation groups in hierarchical structure. A3, A7, A8, A13, A14, A17 represent single annotations in the document. The group named "Content" is the top-level group, other groups "Jane," "Nick," and "Things to confirm" are instances of sub-groups.

The average number of groups created per participant was 8.3 (sd=2.9, min=5, max=14). The average number of annotations per group was 8.2 (sd=2.6, min=1, max=22). Overall, group structuring was not very complex; 45% of the groups had a *flat structure* in that they had no sub-groups nor were they within super-groups, and the remaining 55% of the groups were in *hierarchical structures* (i.e., they had at least one sub-group or a super-group). We analyzed the complexity of these hierarchies in terms of height (the length of the path from the top-level group to the furthest sub-group). Figure 4.3 shows a sample annotation group with a hierarchical structure; the structure has height 2, which is the length of the path from the top-level group named "Content" to the furthest sub-group named "Things to Confirm." We found that the average height of *hierarchical structures* created by participants was only 1.4 (sd=0.5, min=1, max=2).

Analysis on the sub-groups revealed that 44% of them were true sub-components of their higher-level groups (what we call a "proper hierarchy"), e.g., a sub-group named "Missing standard deviation" nested within a group "Missing Information" (since standard deviation is one of the information types presented in the document). The rest of the sub-groups did not reflect proper subset relationships; the hierarchy seemed to result from the individual participant's decision about the relative importance of attributes (what we call an "arbitrary hierarchy"). For example, one participant created a group "For Jane" with a nested sub-group "Clarifications" because she wanted to emphasize and make the recipient-based information more salient. At the same time, another participant created the reversed structure for these same annotations: "Clarifications" with a nested sub-group "For Jane," in which the problem-based information was emphasized.

4.3.4 Strong Support for Structuring

Seven out of eight participants strongly agreed that they liked being able to organize annotations within a document. One of the participants commented that groups were *"infinitely easier than the current annotation format and [can be used] to delegate sections to different authors [which] reduces duplicate effort."* Another participant mentioned that she would like to use annotation groups not only to facilitate her co-authors' workflow but also to manage her own workflow.

4.4 Summary and Additional Comments

All participants structured annotations during the tasks, and perceived the benefits of supporting grouping in annotation tools. We found that participants used different temporal patterns to organize annotations. We identified common semantics of annotation groups, but we did not observe very complex group structuring. It could be that the number of annotations and groups were not large enough to call for complex structures. During the debriefing section, two participants commented that if the number of an-

notations in a group became large (more than about 50 annotations), they would likely have broken their groups down into smaller, more manageable sub-groups. One of the participants who created only "flat" groups said that he did not create nested hierarchy because he felt that the total number of groups he created was manageable, and had he had more groups in the document, he might have created a higher-level group that had some annotation groups as nested sub-groups. More research with larger documents and more iterative cycles of reviewing would be needed to assess the usefulness of supporting complex structuring.

The document topic appeared to have had an impact on the results, which in retrospect was not surprising. The three participants from Psychology created eight or more groups during the task while other participants from Computer Science and Zoology created fewer groups. The most likely explanation for this is derived from the social psychology literature: large numbers of categories reflected users' familiarity with the subject of the document and their thorough understanding of the subject [33]. The structure of some groups resembled a "divide and conquer" problem decomposition approach in which annotation groups and sub-groups correspond to components and subcomponents of the work remaining to be done in the document.

Based on the results from our Paper Prototype Study and the Observational Study (described in Chapter 3), we saw motivations for both structuring and tagging of annotations. We discuss these implications and motivations in the next chapter.

Chapter 5

Incorporating Tagging into Structured Annotations

The Observational Study and the Paper Prototype Study described in the two previous chapters revealed that the ability to organize annotations into groups could benefit co-authors in several ways, namely by facilitating communication, problem decomposition, and workflow management among annotators and co-authors. In this chapter we discuss the implications and motivations for structuring and tagging that were derived from the two studies.

5.1 Implications and Motivations for Grouping and Tagging

The different temporal patterns observed in the Paper Prototype Study for organizing annotations seem to reflect the top-down and bottom-up approaches, as suggested by the information processing literature [18]. What we refer to as top-down occurs when a user creates a group, often a priori, and then adds annotations to that group. Bottom-up occurs when the user formulates the group, often after reviewing all the annotations. We speculate that the *pre-review* approach is a form of *top-down* information processing [18] while *post-review* is similar to *bottom-up*. Reshuffling of annotation groups can be considered as *middle-out* processing. To support these different approaches, it should be possible to create annotation groups at any time, i.e., before, during, or after single annotations are created. We believe that mechanisms for creating and managing annotation groups

should be flexible and light-weight.

Consistent with findings reported by Neuwirth et al. [24], we found that some participants occasionally made *mental notes* about annotations that they wanted to return to for organizing. Hence, providing a light-weight means to externalize such mental notes and support for navigational aids should reduce users' cognitive load. Moreover, tagging-like behaviors found in our Observational Study suggested that tags could also be useful in pinpointing problems in the document.

We saw diversity in the degree of group structuring (proper and arbitrary hierarchies, and flat structures). Disagreement or conflicts among users in defining structures and hierarchies of annotation groups may cause ambiguities and inefficiencies. We argue that tagging is likely a good solution to this issue. With tagging, users do not need to agree on a particular hierarchy, instead they just need to have a shared understanding of a tag's meaning [22]. Hierarchies should still be supported, however, to recognize proper subset relationships among annotation groups.

Although we did not observe very complex group structuring, we imagine that complex structures might arise as the amount of annotations or the size of a collaborative artifact grows over time. We realize that having complex structures might hinder the co-authoring process, because of the additional navigation time required to reach a highly nested annotation and the additional effort required to develop agreement among collaborators about the hierarchical information. This affirms the importance of reducing complexity in the degree of structuring. We note that tagging can be a good approach because it can allow for implicit structures.

5.2 Integrating Structuring and Tagging into Annotations

Based on the implications above, we extended Zheng et al.'s previous structured annotation model [43] by adding tags as one of the optional attributes of annotations (see Table 5.1). Tags in our model serve multifold purposes.

Mandatory Attributes:
Annotator - creator
Timestamp - creation time
Reviewing Status - unread/read and accepted/rejected
Anchor - location in the document
Optional Attributes:
Name - text description
Recipient - intended co-author
Comment - free-form text note
Modification - insert/delete/replace of text
Substructure - list of other associated annotations
Tag - keywords

Table 5.1: Annotation model
from [43] with our added Tag attribute.

They:

- efficiently associate a keyword with annotations,
- externalize mental notes or act as navigational aids,
- easily identify semantic concepts inherent in annotations,
- facilitate workflow by allowing for bottom-up annotation grouping, and
- simplify structures of annotation groups through implicit grouping.

Tags are treated as meta-information about annotations that users can easily add as they review and annotate a document. Tags can be used as navigational aids; by filtering annotations on a particular tag, users can then jump easily between annotations in the selected set of annotations in the document. Users can also choose to save tag-based filter results as structured groups of annotations. Tags allow flexible classification of annotations based on their semantic concepts. Tags provide implicit groups for annotations because co-authors can easily see relationships among annotations labeled

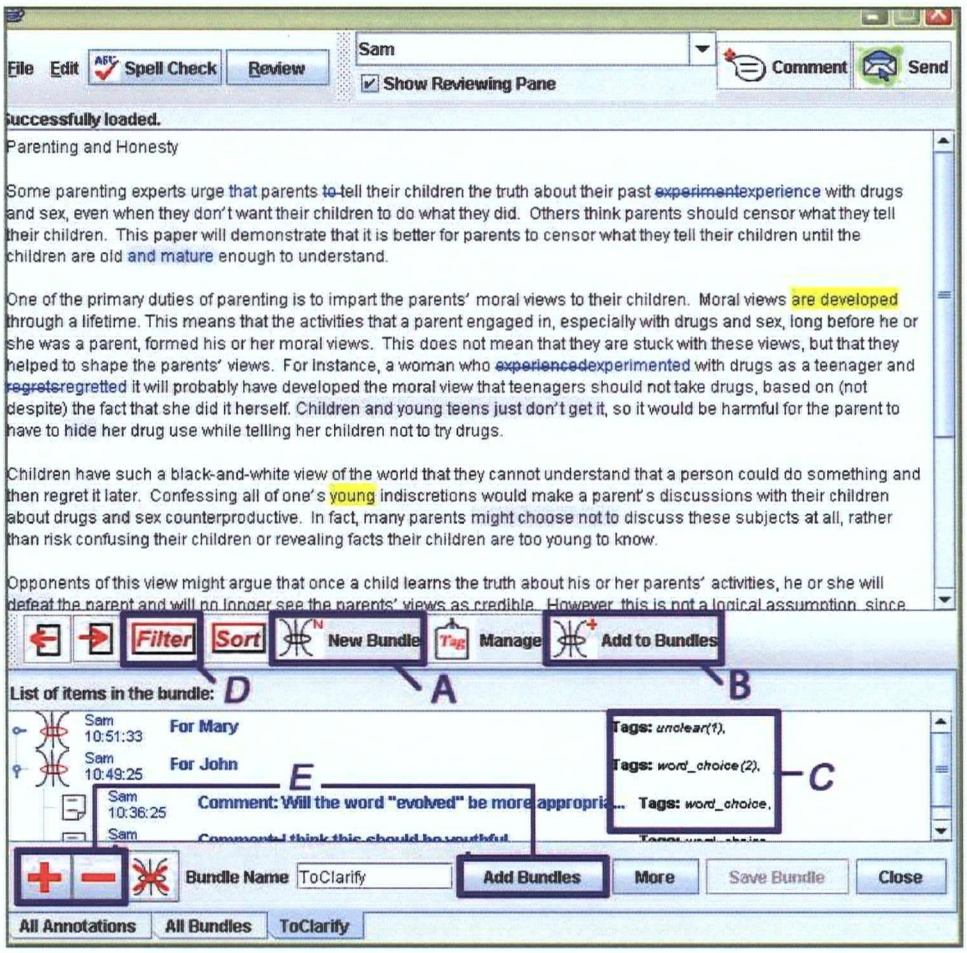


Figure 5.1: Bundle Editor with document and reviewing panes.

with the same tags, even though the annotations are not explicitly grouped together.

5.3 Extended Bundle Editor

We implemented an extended version of *Bundle Editor*, a prototype originally developed by Zheng et al. with structured annotation functions [43]. The main interface to the Bundle Editor prototype (Figure 5.1) consists

of a document pane and a reviewing pane. The document pane serves as a document editor with basic functionalities such as insert, delete, comment, etc. The reviewing pane consists of multiple tabs, each of which displays a specific "bundle" or group of annotations. For detailed information about the Bundle Editor's basic functionalities, refer to Zheng's Master's thesis [42].

Our extensions to the Bundle Editor focus on different ways in which users can create bundles rather than on automatic bundle creation techniques because we believe that automation cannot fully capture the richness and complexity of the annotations used in discussions and workflow management.

The Bundle Editor facilitates multiple approaches to creating and managing the types of bundles that were previously described. The tool supports *top-down* grouping of annotations by allowing users to create a bundle (Figure 5.1A) and then explicitly select annotations to be grouped into that or other bundles (Figure 5.1B). *Bottom-up* grouping is supported by allowing annotations to be tagged with one or more keywords (Figure 5.1C), and then filtered or selected based on their tags (Figure 5.1D, Figure 5.2). *Middle-out* grouping is supported by allowing bundles and annotations to be easily added to or removed from existing bundles (Figure 5.1E).

In order to achieve all the advantages of tags, as described above, we made tagging pervasive throughout the system. Users can easily associate an annotation with one or more tags simply by typing new tags into a textbox or selecting from a list of existing tags (Figure 5.3). Figure 5.2 shows the dialog box with which users can filter annotations based on AND/OR combinations of tags and other attributes such as annotator and annotation type. In the filter dialog box, the prevalence of a given tag is visible to users through the display of its frequency right next to the tag word.

With this implementation of an extended version of the Bundle Editor that robustly supports basic document editing, annotating and structuring of annotations, we proceeded to conduct our final study, in which participants used our prototype for annotating documents and communicating document-related issues.

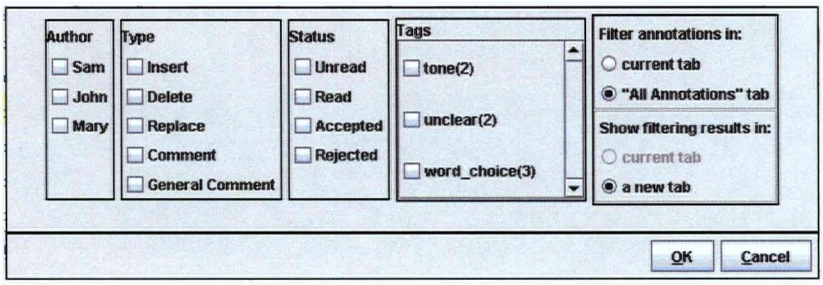


Figure 5.2: Filtering function that allows users to filter annotations based on AND/OR combinations.

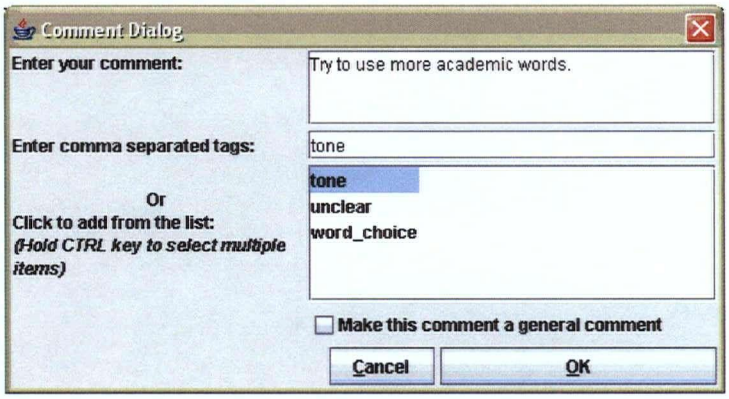


Figure 5.3: Comment dialog box where the user can type in an annotation's comment text and tags.

Chapter 6

Experiment to Examine the Impact of Structuring Annotations

6.1 Study Goals

The Paper Prototype Study described in Chapter 4 showed users' strong support for structuring annotations in a collaborative document. However, important questions regarding the cost-benefit tradeoff associated with structured annotations, and the impact of structuring on the amount of information communicated still remained unadressed. We conducted a controlled experiment and compared the use of structured annotation support to unstructured annotation support for annotating documents and communicating document-related issues. We wanted to examine: (1) if under a controlled comparison with no support for structure users would still perceive the benefits of structuring their annotations (as they had with the Paper Prototype Study); (2) if the overall workload for structuring annotations would be similar to users providing non-structured annotations and then having to communicate any meta information in a text email format (i.e., outside of the document); and (3) if the overall amount of information communicated would differ with structured annotations.

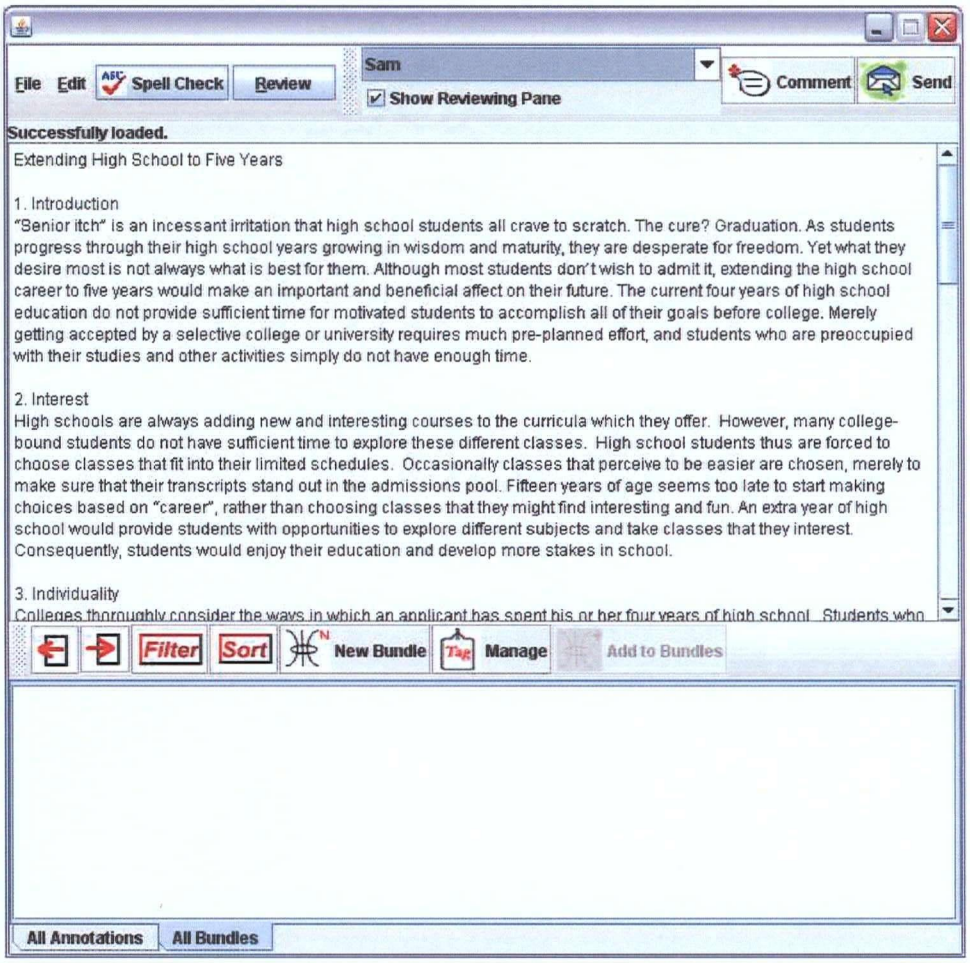


Figure 6.1: The Bundle Editor used in the Experiment. Participants were able to perform basic editing and annotating tasks. The reviewing panel has two tabs by default: "All Annotations" that displays a list of all single annotations, and "All Bundles" that displays a list of all bundles embedded in the document.

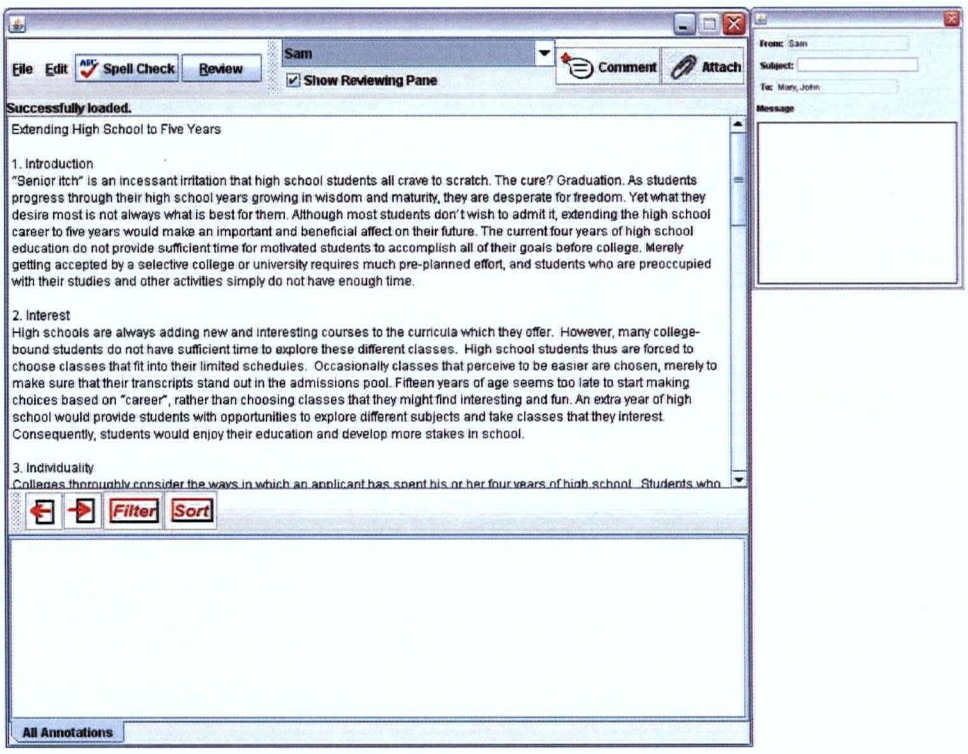


Figure 6.2: The Simple Editor used in the Experiment. Similar to the Bundle Editor, participants were able to perform basic editing and annotating tasks. However, the Simple Editor's reviewing panel has only one tab: "All Annotations" that displays a list of all single annotations embedded in the document. A simulated email message window was also provided in the Simple system condition (shown at the right), allowing for additional document-related communication directed to hypothetical recipients.

6.2 Methodology

We compared two annotation systems: a Bundle system which consists of our extended Bundle Editor (Figure 6.1) and a Simple system which consists of a Simple Editor (an equivalent system that did not support annotation structure), and an email message window (Figure 6.2). Table 6.1 summarizes the differences between the two systems. So as not to bias our participants' perceptions by labeling one as "simple", the two systems were referred to

	Bundle System	**Simple System**
Interface Components	Document panel, *multi-tabbed* reviewing panel	Document panel, *single pane* reviewing panel
Communication Support	Single annotations with *optional user-defined tags, general comments, structured annotations* embedded in the document and listed in the reviewing panel.	Single annotations embedded in the document and listed in the reviewing panel, *a simulated email message window.*
Filtering Functions	AND/OR filtering on Author, Type and *Tag*	AND/OR filtering on Author and Type

Table 6.1: Comparison of Bundle system and Simple system. Both systems were created by modifying our Bundle Editor so that they differed only in the way they supported annotations.

by the experimenter as the Bundle system and the Filter system.

6.2.1 Participants

As in the Observational Study, writing tutors were recruited through online mailing lists and newsgroups. A total of 13 people (all females[1]) participated in the study. They were paid $80 for their time. One of the participants had never used a Microsoft Windows system. Her medical condition (arthritis) made it difficult for her to use a two-button mouse or a keyboard. For this reason, her data was not included in any of the analyses. Data from the other 12 participants were examined and analyzed. None of the participants had previously been involved in the two earlier studies.

In terms of background, all twelve participants used a word processor regularly (mainly Microsoft Word), although six had never used any of its annotation functions. Eleven participants used a word processor frequently (four participants everyday and seven every two to three days) and felt very confident in their usage. The remaining participant only used a word pro-

[1]The population that we recruited from consisted mostly of females, and only females responded to our call for participation.

cessor once a month and her confidence was relatively low. All participants had reviewed documents more than ten times. Seven participants had previously been involved in collaborative authoring, three more than five times, and four fewer than five times.

6.2.2 Task

Participants were asked to review and annotate two documents, one with each of the systems. They were asked to assume a role as a collaborator within a group of three co-authors. The given documents were assumed to be drafted by the other two co-authors with some sections explicitly noted as being jointly drafted by the two, and others separately by each.

Unlike in our Paper Prototype Study, here participants were expected to create annotations, both comments as well as direct edit changes to the document text. They were also requested to provide: (1) their general impression of the writing, and (2) a brief summary of their review that indicated the overall status of the document to help their co-authors skim the document quickly and prioritize the remaining work. The requested feedback was representative of common meta-comments communicated between annotators and recipients, as found in our Observational Study.

The task documents were based on sample essays for the writing sections of the GRE (General Record Examination)[2] and ACT (American College Testing)[3] tests; the contents were on topics of general interest. As shown in Appendix C.3.2, neither of the task documents included any annotations. The documents were manipulated to be isomorphic; they had the same number of problems planted at similar locations throughout the documents. Both syntax problems (e.g., grammar errors, incorrect verb usages) and semantic problems (e.g., unclear statements, unsupported arguments) were carefully planted with the expectation that participants would create both local (sentence-level) and global (document-level) annotations.

[2]GRE is a standardized test for graduate school admissions in North America. The sample GRE essays were collected from http://www.west.net/~stewart/ws/.

[3]ACT is a standardized test for undergraduate college admissions. The sample ACT essays were collected from http://www.actstudent.org/writing/sample/index.html

Categories of planted problems		Num of instances
Semantic	Unclear statements	6
	Flawed logic	3
	Title inconsistency	3
	Unsupported argument	2
	Subtotal:	**14**
Syntax/lexical	Inconsistent wording	2
	Incorrect use of active-passive verbs	4
	Incorrect use of verb tense	5
	Use of contractions	2
	Inappropriate word choice	2
	Subtotal:	**19**
Total:		33

Table 6.2: Problems planted in each task document.
The nine categories are illustrated in Appendix C.3.3.

Prior to the experiment, to test the manipulations, two independent raters with expertise in writing and reviewing documents were asked to identify problems in the documents. They independently identified 72% and 70% of the planted problems across both documents, respectively, with a 60% overlap in the problems identified. The problems that were not identified by either of the two raters were removed from the task documents. In the end, each document had a total of 33 planted problems (as shown in Table 6.2). Appendix C.3.3 shows these problems in the context of the study documents. A third document was used during practice sessions. Because the same practice document was used in every configuration, we did not control the number or the types of problems in the practice document.

6.2.3 Procedure

Each participant had a single four-hour session. It began with a demographic questionnaire to obtain past computer, co-authoring and reviewing experience. Participants then saw a training video on general concepts such as collaborative writing and annotations, and how to use their first assigned

system. To ensure that all participants would have a similar level of familiarity with a system's functionality, participants had 15 minutes to perform a set of guided annotation tasks to gain experience with that system. They were provided with a list of annotations to create in the practice document. After the 15-minute practice session, participants had one hour to perform the experimental task on their first document with their first system. A questionnaire followed to collect feedback on that system. Participants were given a short 10 minute break and were then shown a training video on how to use their second system, followed by a 15-minute practice session, then the experimental task on their second document with their second system. Another questionnaire was conducted to gather feedback on that system. To solicit comparative feedback between the two systems, a final questionnaire was administered. We will refer to this final questionnaire as the *comparative questionnaire* and the previous two questionnaires as *system-specific questionnaires.* Appendix C.2 shows all the questionnaires used in the study. A short semi-structured interview to collect further information regarding preferences and perceived performance ended the sessions.

6.2.4 Design

The experiment used a within-subjects (system type) factorial design. Document was a within-subjects control variable, and both system and document presentation orders were between-subject controls. To minimize learning effects, we counterbalanced the order of presentation for both system type and document, resulting in four configurations.

6.2.5 Measures

The amount of communication among co-authors was assessed by counting the number of (1) single annotations and (2) meta-comments. Single annotations were the same in both systems: edits and comments. Counting meta-comments differed between the two systems. In the Bundle system, each unique bundle and tag as well as each general comment counted as one meta-comment. In the Simple system, email content was analyzed to extract

meta-information items; e.g., a statement saying "Try to use more academic words in the places I highlighted" was counted as one meta-comment. To ensure that all meta-comments that were created were counted, any single comments anchored at the beginning or the end of documents in both systems were also analyzed to see if they contained meta-information. Self-reported measures from the two system-specific questionnaires were used to assess subjective workload measures associated with each task using the NASA-TLX workload index[4]. We also included other measures such as performance satisfaction, time satisfaction, reviewing efficiency, reviewing accuracy, and ease of use at the end of each system-specific questionnaire to assess perceived benefits associated with each system. The comparative questionnaire included the same set of questions as in the system specific questionnaires. Each question asked participants to rate the two systems on a single 10-point Likert scale (1:low and 10:high) using different notations (e.g., using 'X' for the Bundle system, and 'O' for the Simple system). During the interview, participants were asked to comment on the cost-benefit tradeoff of using each system.

6.2.6 Hypotheses

Communication Hypotheses: (a) Participants will create more meta-comments in the Bundle system than in the Simple system. (Creating structured annotations is an easier way to provide meta-commentary than doing so separately in the body of an email.) (b) Participants will create similar numbers of single annotations in both systems. (Both systems have identical support for single annotations except for tags, which are a type of single annotation that is only in the Bundle system.)

Workload Hypothesis: Reviewing with the Bundle system will not require significantly higher workload than reviewing with the Simple system. (The added effort to group and tag annotations will not be greater than

[4]NASA-TLX is a standardized multi-dimensional rating procedure that derives an overall workload score based on a weighted average of ratings on six workload categories[15].

that required to compose a detailed email message with the equivalent information.)

Cost-Benefit Hypothesis: Participants will perceive the net gain — the amount by which the benefit exceeds the cost — to be higher in the Bundle system than in the Simple system.

6.3 Results

We report on the quantitative data along with the qualitative data and feedback provided during the interview. Before testing our hypotheses, we checked to make sure that there was no effect of document; we ran a series of 2 documents x 2 orders of systems x 2 orders of documents ANOVA tests on our dependent measures, which showed no significant main or interaction effects of document[5]. Satisfied that there was no effect of document, we then ran a series of 2 systems x 2 orders of systems x 2 orders of documents ANVOA tests to evaluate our hypotheses. Along with statistical significance for each of these ANOVA tests, we report partial eta-squared η^2, a measure of effect size, which is often more informative than statistical significance in applied human-computer interaction research [19]. To interpret this value, 0.01 is a small effect size, 0.06 is medium, and 0.14 is large [9].

6.3.1 Communication

Participants created an average of 64.5 single annotations (sd=27.5) in the Bundle system and 76.1 (sd=40.2) in the Simple system, a difference that was not statistically significant ($F(1,8)=2.20$, $p=0.18$, $\eta^2=0.22$). Figure 6.3 shows the number of single annotations created by participants. Although

[5]Results from testing main effects on our three dependent measures were as follows: number of single annotations, ($F(1,8)=1.12$, $p=0.32$, $\eta^2=0.12$); number of meta-comments, ($F(1,8)=1.46$, $p=0.26$, $\eta^2=0.16$); and TLX workload measure, ($F(1,8)=0.61$, $p=0.46$, $\eta^2=0.07$). While none of these tests showed a significant effect of document, we note that there were large effect sizes for the number of single annotations and meta comments. Given that we fully counterbalanced the order of documents, this is not a big concern for our study, but it would be interesting to examine this further.

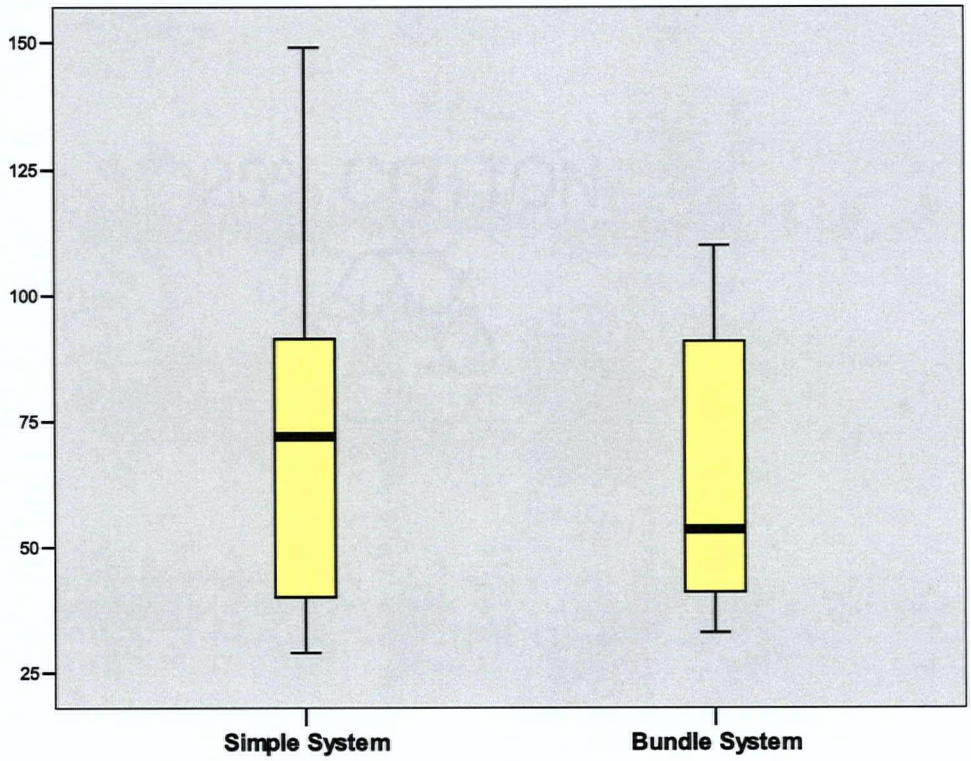


Figure 6.3: Box-plot of the number of single annotations communicated in each system (N=12). The bars show the range of values, the boxes span between the first and third quartiles, the horizontal lines within the boxes indicate the median values.

this lack of effect was expected and supports our communication hypothesis for these types of annotations, we note that a large effect size was found, suggesting that it may be prudent to validate this finding with further research.

In terms of meta-comments, two main categories were observed: (1) recipient-based (e.g., a to-do list for Nick) and (2) problem-based (e.g., tone of the document). The categories were not exclusive; in some cases, the same annotation(s) was (were) counted as both types, e.g., an annotation tagged with the content-related info "argument" and associated with a to-do bundle "For Mary," or a statement saying "Mary, you should watch out for

unsupported claims in paragraph 3."

Participants created significantly more meta-comments ($F(1,8)=13.09$, $p=0.01$, $\eta^2=0.62$) when reviewing with the Bundle system (avg=6.7, sd=3.3) than reviewing with the Simple system (avg=3.7, sd=2.2), also supporting our communication hypothesis. Figure 6.4 shows the number of meta-comments created by participants.

Interestingly, we found a significant system order effect on the number of meta-comments ($F(1,8)=17.44$, $p<0.01$, $\eta^2=0.69$). Participants who were exposed to the Bundle system first included significantly more meta-comments across both systems (avg=7.0, sd=2.4) than those who used the Simple system first (avg=3.3, sd=1.6). One explanation is that the Bundle system facilitated meta-comments in the first reviewing task, leaving participants with the inclination to similarly provide more information in the second task.

Consistent with the quantitative data, many participants also said that they were able to provide a more comprehensive review using the Bundle system. One participant said *"[The Bundle system] maximizes the interaction between the writers,"* another said *"[When] you need a more critical approach [it] gives you the exact tools,"* and another participant added *"I could communicate more information [that] I think is important to get across. ... I was not just correcting the problems; I had a chance to explain why, like to justify it."* These qualitative comments are consistent with our communication hypotheses.

6.3.2 Self-assessed Workload

Perceived average workload with the Bundle system, as measured by the NASA-TLX, was 69.8 (sd=8.0) while that associated with the Simple system was 63.0 (sd=11.8), which was a marginally significant difference ($F(1,8)=4.53$, $p=0.07$ and $\eta^2=0.36$). The NASA-TLX measures workload on an overall scale of 0:low to 100:high. Figure 6.5 shows each individual participant's reported TLX workload. It indicates that eight out of twelve participants associated a higher workload with the Bundle system, two the reverse, and

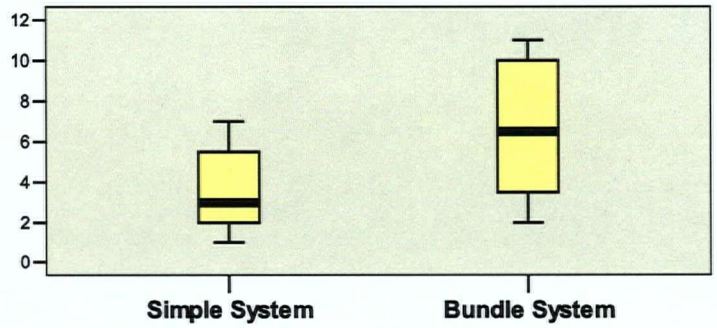


Figure 6.4: Box-plot of number of meta-comments communicated in each system (N=12). The bars show the range of values, the boxes span between the first and third quartiles, the horizontal lines within the boxes indicate the median values.

the remaining two rated workload equally. Finding a difference in workload was not consistent with our hypothesis, but may be explained by the fact that participants included more meta-comment information in the Bundle system, thus requiring more work; alternately, it could be because of usability issues that were uncovered with the Bundle system, as described later in this chapter. More research is required to tease this apart.

Results for each individual workload category and benefit question are displayed in Table 6.3, as analyzed using the Wilcoxon Signed-Rank test[6]. On average, the Bundle system required significantly higher mental demand (p=0.01) and more effort (p=0.03) than the Simple system. The average temporal demand associated with the Bundle system was marginally significantly higher than that associated with the Simple system (p=0.07). The average ratings for performance and process satisfaction associated with the Simple system were statistically significantly higher than those associated with the Bundle system (p=0.04 and p=0.01 respectively). Participants also felt that the Simple system was easier to use than the Bundle system (p=0.02). These findings may be explained by the fact that the features and the interaction techniques required in the Bundle system were more so-

[6] Neither a t-test nor ANOVA was used because data was not normally distributed.

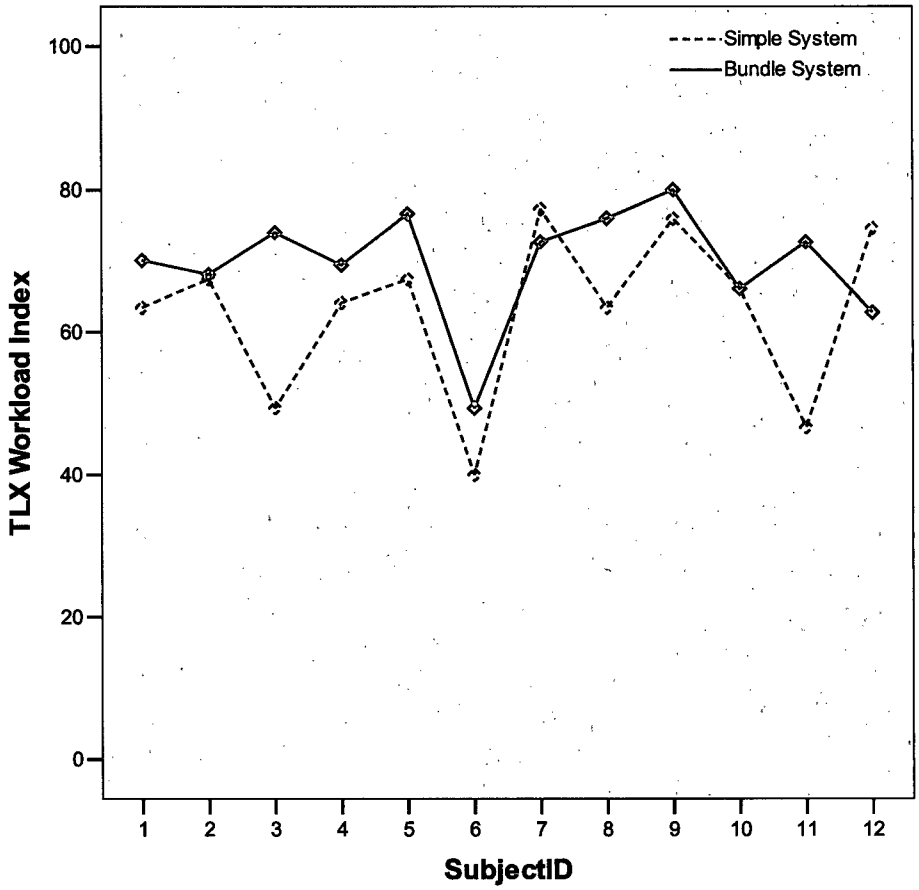


Figure 6.5: Overall workload associated with each system (N=12).

phisticated than those in the Simple system; hence, some participants might have needed more training and practice than they were given in the study, and felt less comfortable with the Bundle system.

More research is needed to improve the usability of the Bundle system, and we should examine the workload associated with the system over an extended period of use.

	Mean		Std. Deviation		Sig.
	Simple	*Bundle*	*Simple*	*Bundle*	
Mental Demand	6.33	8.25	1.97	1.77	.01*
Physical Demand	3.67	5.00	1.97	2.70	.12
Temporal Demand	4.25	5.25	2.34	2.73	.07
Performance	7.83	7.08	1.59	1.56	.04*
Effort	6.08	7.33	2.11	1.83	.03*
Frustration	4.17	5.00	2.79	2.52	.40
Process Satisfaction	8.58	6.83	1.08	2.08	.01*
Time Satisfaction	8.50	8.17	1.24	1.47	.49
Reviewing Efficiency	7.92	7.25	1.38	2.26	.55
Reviewing Accuracy	8.08	7.50	1.08	2.32	.91
Ease of Use	8.08	6.50	2.43	2.47	.02*

Table 6.3: Mean responses to self-reported measures with the scale 1:low to 10:high. A * indicates that the difference between the means are statistically significant with $p<0.05$ (N=12).

6.3.3 Cost-Benefit

During the interview, participants were asked to comment on the cost-benefit tradeoff for using each system. Eleven of twelve participants found both systems to be useful and to have positive net gain (the benefit outweighed the cost). Among these 11 participants, eight said the net gain was higher in the Bundle system than in the Simple system, and that they would definitely like to use the Bundle system in their future annotating tasks. These participants acknowledged that even though the cost of using the Bundle system was higher than for the Simple system (as reflected in our workload measure), the Bundle system would return a much greater benefit, especially over iterative collaborations. For example, one participant explained that because she was able to provide the authors with a more comprehensive review using the Bundle system,

> *"going forward, if I am working with [the same co-authors] again, they would already know what kind of things I am looking for. So it's like I do the front heavy loading [by putting in extra*

Cost (α=0.87)	Mental Demand Physical Demand Frustration
Benefit (α=0.76)	Reviewing Efficiency Reviewing Accuracy

Table 6.4: Factors used for measuring cost and benefit and their Cronbach's alpha values.

> *effort for this first iteration] ... As you front load the work, as you go through [over iterations], it only gets easier."*

The remaining three participants (out of eleven) mentioned that although the Bundle system's benefit was higher than the Simple system, the cost associated with the Bundle system was much higher, resulting in a lower net gain compared to the Simple system. Nonetheless, they said that they would use the added functionality in the Bundle system in some of their future annotating tasks when they needed to provide detailed and precise feedback on larger documents.

The remaining one participant believed that the returned benefit was not worth the cost for either system. When forced to choose between the two systems for her future use, she chose the Simple system because she preferred to give free-form non-structured feedback similar to the verbal feedback to which she was more accustomed. Thus, while the majority of participants thought that structuring annotations was worth the effort, there was clearly some diversity of opinion on this point.

From the comparative questionnaire data, we computed an estimate of the participants' perceived cost and benefit associated with one system relative to the other. We used an average of the responses to the Mental Demand, Physical Demand, and Frustration from the TLX index as a measure of cost, and the Reviewing Efficiency and Reviewing Accuracy as a measure of benefit (see Table 6.4). We did not include the other dimensions, Temporal Demand, Performance, Effort, Process Satisfaction, Time Satisfaction and Ease of Use in the calculations because of their lack of consistency with

other measures, as was evident from low Cronbach's alpha values[7] (α), a reliability measure.

As shown in Figure 6.6, perceived cost associated with the Bundle system was 7.4 (sd=1.8) while that associated with the Simple system was 5.0 (sd=2.2), a significant difference ($t(11)=6.06$, $p<0.00$) according to a paired t-test. Perceived benefit associated with the Bundle system was 7.9 (sd=1.6) while that associated with the Simple system was 7.8 (sd=1.2), a non-significant difference ($t(11)=6.06$, $p=0.93$). Thus, the benefit measures were similar in the two systems but the cost measure was higher in the Bundle system than in the Simple system.

At first glance, these quantitative results from the comparative questionnaire seem inconsistent with the qualitative feedback obtained during the interview. The most probable explanation, however, is that the quantitative questionnaire responses reflected the cost-benefit tradeoff with respect to the study tasks that participants had just finished for the given scenario, while the qualitative interview reflected the tradeoff with respect to participants' envisioned future use of an annotation system (on larger documents and over iterations). Clearly, additional research will be necessary to probe the cost-benefit tradeoff associated with the extended use of the Bundle system over time.

6.3.4 Summary of Results

We summarize our results for each of our hypotheses.

Communication Hypotheses: these were supported. Participants created more meta-comments in the Bundle system than in the Simple system, and similar numbers of single annotations in both systems.

[7] Cronbach's alpha measures how well a set of items (or variables) measures a single uni-dimensional construct. When data have a multi-dimensional structure, Cronbach's alpha will usually be low. To interpret Cronbach's alpha values, 0.7 is considered to be an acceptable reliability coefficient although lower thresholds are sometimes used in the literature [32].

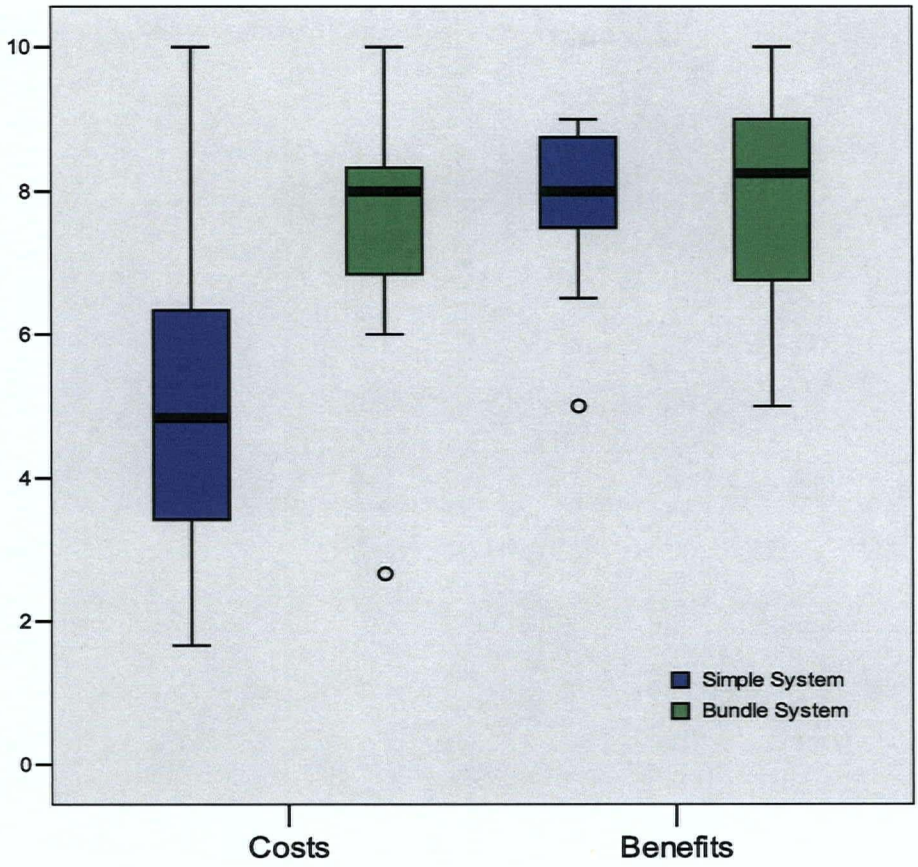


Figure 6.6: Box-plot of costs and benefits measures for each system (N=12). The bars show the range of values, the boxes span between the first and third quartiles, the horizontal lines within the boxes indicate the median values. A circle identifies an outlier.

Workload Hypothesis: this was not supported. Reviewing with the Bundle system showed significantly higher workload than reviewing with the Simple system.

Cost-Benefit Hypothesis: this was partially supported. According to the qualitative feedback, the majority of participants thought that the Bundle system offered higher net gain, while the quantitative results suggested

otherwise.

6.3.5 Other Measures: Usage of Structures

Bundles: Eleven out of twelve participants created bundles. Analysis of the bundles[8] and their associated annotations revealed that bundles were used to communicate both problem-based (59%) and recipient-based (41%) types of meta-information.

Problem-based bundles were created using four different approaches: (1) unique-tag based, (2) clustered-tags based, (3) tag-less, and (4) type based. In the unique-tag based approach, bundles were created by grouping annotations based on a particular tag. For example, a bundle named "word choices" was created by manually selecting annotations tagged with "word choices" or by saving a filter result on the tag "word choices." In the clustered-tag based approach, bundles were created by manually selecting and aggregating all or a subset of annotations that were labeled with similar tags. For example, a bundle named "Argument" was created by grouping annotations that were labeled with tags named "argument" and "unclear." In the tag-less approach, bundles were created by manually selecting and grouping annotations that did not have any tags. Lastly, in the type based approach, a bundle was created by saving a filter result on a particular type of annotation, such as "General Comment." Figure 6.7 shows the number of bundles created using each approach. The majority of problem-based bundles were created based on a unique tag or a set of similar tags, using the "bottom-up" structuring approach previously described. The majority of participants mentioned that creating bundles using this approach was easier and less time-consuming than the top-down approach, which was used to create recipient-based bundles.

Recipient-based bundles were created by manually selecting and grouping annotations that had diverse sets of associated tags. We believe that these recipient-based bundles were created to help the hypothetical recipients manage their workflows. This was explained explicitly by one partici-

[8]One empty bundle, i.e., a bundle that did not have any associated annotations, was excluded from the analysis.

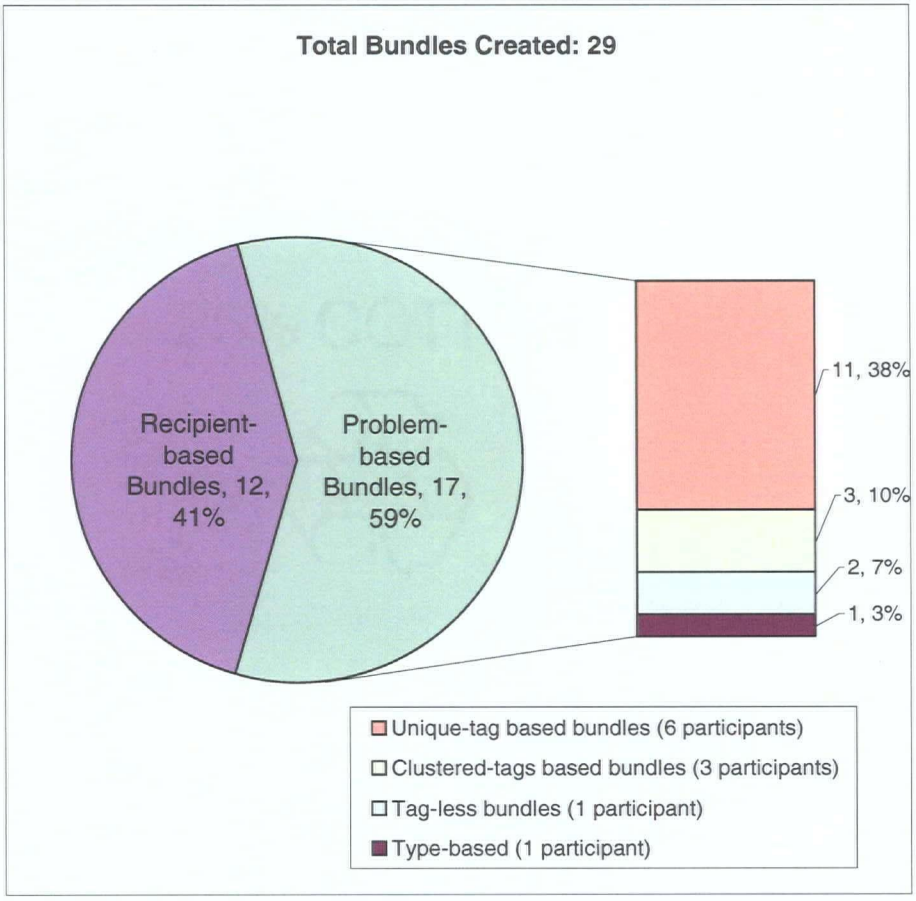


Figure 6.7: All bundles created by 11 participants.

pant, who reflected on one of her previous collaborative writing experiences and stated that:

> *"Bundles would have been useful for addressing the co-authors' problems individually. They didn't have the same ... errors. So, being able to separate them out, say you need to work on this, you need to work on that. But then also for the things that they were working on together, [bundles would allow me] to be able to combine them as well. So it's a way of both separating them out*

but then making [them] more cohesive at the same time."

Another participant explained that workflow-related information communicated through bundles could help recipients review annotations efficiently because *"everything that requires a certain way of dealing with is together [in bundles]"* and hence, co-authors *"don't have to keep switching their mindset from one thing to another."*

Tags: All twelve participants added tags to their annotations. Tags were used to describe specific problem(s) of the annotations that they were added to. For example, one annotation that pointed out an argument problem as "This brings in a whole separate issue that you're not prepared to deal with in this paper," had a tag named "argument."

It was surprising that participants did not use tags at all to communicate recipient-based meta-comments; they used tags exclusively to communicate problem-based kinds of information. Participants said that the information communicated through tags brought *"awareness to patterns of problems in documents,"* and would allow co-authors to see the strengths and weaknesses of their writing, and also to achieve a quick overview of the current document status. One participant further stated that the information in tags allowed her to achieve *"a greater perspective on the reviewing process"* that she had gone through. We found that tags were sometimes used as alternatives to long comments when addressing recurring problems. This was explained explicitly by one participant during the interview, *"One can comment the first time one runs into a problem. But after that, [tags] are like reminders, almost to go back to that comment."*

Relationship between bundles and tags: Table 6.5 provides a high-level description of the bundles and tags that participants created. On average, each participant grouped 31% of her annotations into one or more bundles, and labeled 23% of the annotations with one or more tags; 10% of the annotations were both tagged and bundled, while 56% were neither tagged nor bundled. The average numbers of bundles and unique tags created per participant were 2.4 (sd=1.4), and 4.8 (sd=2.8), respectively. While

	# of bundles	# of anno in each bundle	% of anno bundled	# of unique tags	# of anno with each tag	% of anno tagged
Average	2.42	7.68	30.85	4.75	3.32	23.01
Median	2.00	5.50	13.70	4.00	3.09	17.17
Std Dev	1.44	5.61	33.14	2.80	1.85	18.94
Min	0.0	0.0	0.0	1.0	1.14	9.09
Max	5.0	19.4	100.0	11.0	7.5	81.36

Table 6.5: Descriptive summary of bundles and tags (N=12).

a greater number of unique tags were created than bundles, a higher percentage of annotations were associated with bundles than with tags; the number of annotations per bundle (avg=7.7, sd=5.6) was higher than that per tag (avg=3.3, sd=1.9). This may be explained by the fact that tags were used more to label comments than edits, while bundles were used almost equally to organize both comments and edits. On average, participants tended to create more edit annotations than comment annotations; 71% of all annotations were edits in each document.

In order to better understand how bundles and tags were related, bundles are displayed in Figure 6.8 as a function of the number of unique inherited tags (tags attached to the grouped annotations in a bundle) and the number of annotations in each bundle. The number of unique inherited tags was small for most bundles.

6.3.6 Additional Qualitative Feedback

Usability of the Bundle system: Six participants suggested that the Bundle system needed to be more intuitive and straightforward. The interaction technique for adding/removing annotations to/from bundles was a bit cumbersome: a separate tab for each bundle had to be opened and more than one button clicks were required to add/remove each annotation. Improving the usability of the system would involve implementing more efficient

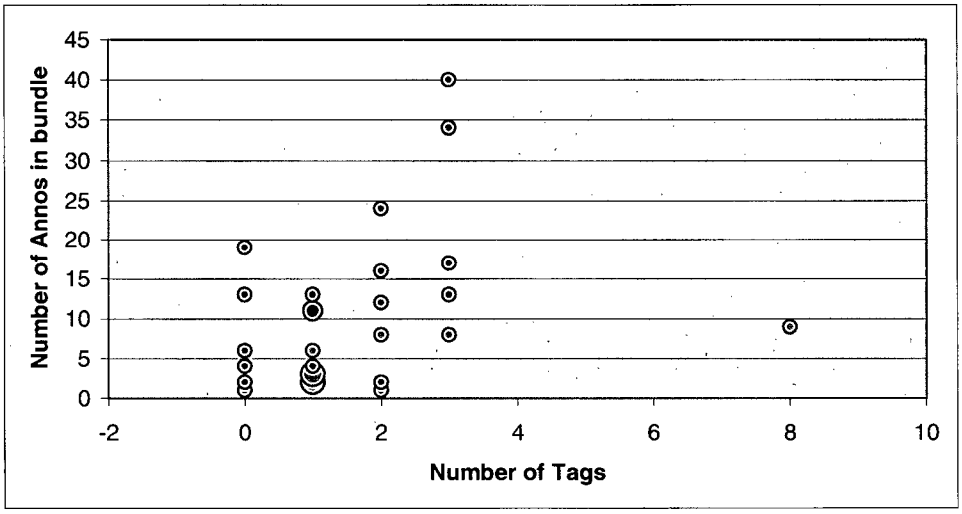


Figure 6.8: Bundles as a function of number of unique inherited tags and number of annotations; the size of each bubble shows the number of bundles at each coordinate.

interaction techniques for annotation organization, such as drag-and-drop.

Preference for Structuring: Four of the six participants who were exposed to the Bundle system first said that while performing the second task with the Simple system, they wished it had some of the Bundle system functionality (e.g., tagging or grouping annotations). They felt that they could not provide feedback *"as precise and thorough as in the Bundle system"* and they had difficulty *"explaining how problems [were] connected and uniting comments."*

6.4 Discussion

Structured annotations are worth the effort: Almost all participants (11/12) believed that structuring annotations offered *positive* net gain; structured annotations are worth the effort. Furthermore, the majority of participants (8/12) stated a preference to use structured annotations in their future

work because they believed that structuring annotations offered *higher* net gain relative to simple annotations. For the remaining participants, the perceived benefits did not sufficiently outweigh the additional workload associated with structured annotations; nevertheless they said that they would use structured annotations in certain contexts where they needed to annotate documents thoroughly.

Top-down and bottom-up approaches complement each other: Although the bottom-up approach was considered to be less time consuming and more light-weight, it did not displace the top-down approach. Both approaches were used by almost all participants (11/12). The one participant who did not use bundles commented during the interview that she ran out of time to create bundles towards the end of the task. We suspect that a top-down grouping approach was used to create recipient-based bundles because the information regarding intended co-author(s) was known prior to annotating the documents. Conversely, a bottom-up approach was used to create problem-based bundles probably because the structure was formulated only after realizing relationships among annotations. Hence the top-down and bottom-up approaches support structuring annotations in a complementary way, and both should be supported in an annotation-structuring tool.

Structured annotations promote valuable communication: Participants created more meta-comments in the Bundle system than in the Simple system. Participants' remarks on the usefulness of bundles and tags as described earlier suggest that the meta-information communicated in the form of structured annotations was perceived to be valuable and beneficial for both the intended recipients and the annotators themselves, by allowing for a comprehensive review that went beyond simple annotations, providing for a greater perspective on the reviewing process, supporting workflow management, and offering a quick overview of the document status.

Chapter 7

Conclusions and Future Work

We assessed the impact of supporting structured annotations on users who create annotations. The Observational Study strengthened our understanding of how annotators use existing tools to communicate document-related information in the form of annotations. That study also revealed annotation practices that could benefit from additional structuring support, such as annotators describing how annotations in a document relate to each other. In the Paper Prototype Study, we investigated how annotators would structure annotations, if given the option. Common behaviour that emerged was the grouping of thematically related annotations, as well as the grouping of annotations specifically targeted to a given co-author. The study also uncovered a range of temporal approaches to structuring annotations, such as top-down and bottom-up grouping. In the Experiment, we compared the use of structured annotation support to unstructured annotation support for annotating documents and communicating document-related issues. We found that structured annotations were perceived to be worth the effort despite the additional workload, and that the tag-based grouping approach that we added to an existing annotation model [43] complements the original hierarchical approach in describing relationships among annotations in the document.

7.1 Limitations

In all three of our studies, all participants were provided with documents to perform the tasks. In order to keep the tasks manageable,we chose relatively

simple documents in terms of length compared to the documents that our system would likely be used for in actual practice. The generalizability of our results to larger documents is thus somewhat limited. Moreover, we carried out our research as laboratory studies. As with any lab study, there is a trade-off between realism and generalizability for increased precision [23]. We tried to address the limitations and maintain a degree of ecological validity by designing the study scenarios to be representative of those commonly found in co-authoring projects, and designing the task documents to have the types of errors common in writing.

Another limitation lies in our choice of involving professional writing tutors as participants in the Observational Study and the Experiment. We recruited professional writing tutors as participants because of their experience in annotating documents, as described in the Chapter 1. However, we realize that tutors are not necessarily identical in their work practices to small groups of co-authors. Further investigation may be required to better understand the generalizability of our results to co-authors, or to other user groups.

7.2 Future Work

We envision our structured annotation model as being integrated into existing word processors (such as Microsoft Word or Google Docs) to support document-embedded communication among co-authors. A few avenues for future research are summarized in this section.

7.2.1 Streamlining the Additional Workload

Our findings from the Experiment indicate that creating structured annotations increases workload. It is possible that with more exposure and an extended period of use, the workload associated with the Bundle system might decrease. Nevertheless, further work is needed to explore how the additional workload can be streamlined. As discussed earlier, improving the usability of the structuring tool should decrease workload. Hence the

usability issues described earlier need to be resolved; efficient and intuitive interaction techniques for grouping annotations need to be incorporated.

7.2.2 Evaluating Structured Annotations in Iterative Collaboration

Zheng et al. assessed the impact of structured annotations on users who receive structured annotations [43]. We have now assessed the impact on users who create them. The next step will be to investigate the impact of structured annotations on the complete co-authoring workflow that involves iterative cycles of annotating and editing. More specific issues to explore include how users might go about managing and maintaining structured annotations as the document evolves with iterations of the reviewing cycle.

7.2.3 Incorporating Structured Annotations into E-mail

The purpose of structured annotations is to unify all document-related communication into the document. The purpose is not to completely displace external communication methods such as email. Given its pervasiveness, communication through email is inevitable. Perhaps duplicating or summarizing information from structured annotations in the body of email messages might appropriately serve as a detailed notification; this could allow co-authors to get a quick overview of the reviewed document and the embedded annotations without having to open the document. A major drawback to any use of e-mail is the seemingly inevitable urge users have to send a reply, which runs the risk of having document-related communication outside the document. More research is definitely needed to explore ways in which email content can be generated automatically from the information communicated in structured annotations in a document and whether it is possible to automatically incorporate e-mail replies explicitly back into the document.

Bibliography

[1] Del.icio.us [online social bookmarking tool], 2007. Accessed May 2, 2007, from `http://del.icio.us`.

[2] Flickr [online photo management and sharing tool], 2007. Accessed May 2, 2007, from `http://flickr.com`.

[3] Google docs [online word processor tool], 2007. Accessed May 2, 2007, from `http://docs.google.com`.

[4] Write board [online collaborative writing tool], 2007. Accessed May 2, 2007, from `http://www.writeboard.com`.

[5] D. Abrams, R. Baecker, and M. Chignell. Information archiving with bookmarks: Personal web space construction and organization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 41–48. ACM Press, 1998.

[6] R. M. Backer, D. Nastos, I. R. Posner, and K. L. Mawby. The user-centered iterative design of collaborative writing software. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 399–405. ACM Press, 1993.

[7] S. Bloehdorn and M. Volkel. Tagfs — tag semantics for hierarchical file systems. In *Proceedings of International Conference on Knowledge Management*, 2006. Retrieved May 2, 2006, from `http://semfs.ontoware.org/pubs/2006-09-iknow2006-tagfs.pdf`.

[8] E. Churchill, J. Trevor, S. Bly, L. Nelson, and D. Cubranic. Anchored conversations: Chatting in the context of a document. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 454–461. ACM Press, 2000.

[9] J. Cohen. Eta-squared and partial eta-squared in communication science. *Human Communication Research*, 28:473–490, 1973.

[10] L. Ede and A. Lunsford. *Singular Texts/Plural Authors: Perspectives on Collaborative Writing.* Southern Illinois University Press, Carbondale, IL, 1990.

[11] C. A. Ellis, S. J. Gibbs, and G. L. Rein. Groupware: Some issues and experiences. *Communications of the ACM*, 34(1):39–58, 1991.

[12] R. S. Fish, R. E. Kraut, and M. D. Leland. Quilt: a collaborative tool for cooperative writing. In *Proceedings of the ACM SIGOIS and IEEECS TC-OA Conference on Office Information Systems*, pages 30–37. ACM Press, 1988.

[13] S. Golder and B. Huberman. The structure of collaborative tagging systems. *Journal of Information Science*, pages 198–208, 2006.

[14] J. Grudin. Why cscw applications fail: Problems in the design and evaluation of organizational interfaces. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, pages 85–93. ACM Press, 1988.

[15] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Human Mental Workload*, pages 139–183, 1988.

[16] W. Jones, A. J. Phuwanartnurak, R. Gill, and H. Bruce. Don't take my folders away! organizing personal information to get things done. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1505–1508. ACM Press, 2005.

[17] H. Kim and K. S. Eklundh. Reviewing practices in collaborative writing. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, pages 247–259. ACM Press, 2001.

[18] R. A. Kinchla and J. M. Wolfe. The order of visual processing: "top-down", "bottom-up" or "middle-out". *Perception and Psychophysics*, 25(3):225–231, 1979.

[19] T. Landauer. *Chapter 9: Behavorial research methods in human-computer interaction*, pages 203–227. Handbook of Human-Computer Interaction. Elsevier Science, Amsterdam, second edition, 1997.

[20] D. Lawton and I. E. Smith. The knowledge weasel hypermedia annotation system. In *Proceedings of Hypertext and HyperMedia*, pages 106–117. ACM Press, 1993.

[21] P. B. Lowry, C. C. Albrecht, J. D. Lee, and J. F. Nunamaker. Users experiences in collaborative writing using collaboratus, an internet-based collaborative work. In *Proceedings of International Conference on System Sciences*, pages 21–30. IEEE Computer Society, 2002.

[22] A. Mathes. Folksonomies - cooperative classification and communication through shared metadata. Retrieved June 20, 2006, from `http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html`, 2004.

[23] J. McGrath. Methodology matters: Doing research in the behavioural and social sciences. *Readings in Human-Computer Interaction: Towards the Year 2000*, 1994.

[24] C. Neuwirth, R. Chandhok, D. Charney, P. Wojahn, and L. Kim. Distributed collaborative writing: A comparison of spoken and written modalities for reviewing and revising documents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 51–57. ACM Press, 1994.

[25] C. Neuwirth, D. S. Kaufer, R. Chandhok, and J.H. Morris. Issues in the design of computer support for co-authoring and commenting. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, pages 183–195. ACM Press, 1990.

[26] S. Noel and J.M. Robert. Empirical study on collaborative writing: What do co-authors do, use and like? In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, volume 13, pages 63–89. ACM Press, 2004.

[27] I. Ovsiannikov, M. Arbib, and T. Mcneill. Annotation technology. *International Journal of Human-Computer Studies*, 50:329–362, 1999.

[28] J. Porter. Folksonomies: A user-driven approach to organizing content. Retrieved March 20, 2007, from `http://www.uie.com/articles/folksonomies/`, 2006.

[29] I. R. Posner and R. M. Baecker. *How People Write Together*. Morgan Kaufmann, San Mateo, CA, 1993.

[30] D. Quan, K. Bakshi, D. Huynh, and D. Karge. User interfaces for supporting multiple categorization. In *Proceedings of International Conference on Human-Computer Interaction*, 2003.

[31] R. Rimmershaw. Collaborative writing practices and writing support technologies. *Instructional Science*, 21:15–28, 1992.

[32] J. R. Santos. Cronbach's alpha: A tool for assessing the reliability of scales. *Journal of Extension*, 37(2), 1999.

[33] W. A. Scott. Cognitive complexity and cognitive flexibility. *Journal Information for Sociometry*, 25(4):405–414, 1962.

[34] R. Sinha. A cognitive analysis of tagging. *Thoughts on Technology, Design, and Cognition.* Retrieved June 7, 2007, from`http://www.rashmisinha.com/archives/05_09/tagging-cognitive.html`, 2005.

[35] M. Storey, L. Cheng, I. Bull, and P. Rigby. Shared waypoints and social tagging to support collaboration in software development. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, pages 195–198. ACM Press, 2006.

[36] J. Teevan, C. Alvarado, M. S. Ackerman, and D. R. Karger. The perfect search engine is not enough: A study of orienteering behavior in directed search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM Press, 2004.

[37] C. Weng and J. H. Gennari. Asynchronous collaborative writing through annotations. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, pages 564–573. ACM Press, 2004.

[38] C. Weng, D. W. McDonald, and J. H. Gennari. A collaborative clinical trial protocol writing system. *Journal of Medical Informatics*, pages 1461–1465, 2004.

[39] C. Weng, D. W. McDonald, and D. Sparks. Participatory design of a collaborative clinical trial protocol writing system. *International Journal of Medical Informatics*, pages 245–251, 2006.

[40] S. Whittaker and C. Sidner. Email overload: Exploring personal information management of email. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 276–283. ACM Press, 1996.

[41] P. Wojahn, C. Neuwirth, and B. Bullock. Effects of interfaces for annotation on communication in a collaborative task. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 456–463. ACM Press, 1998.

[42] Q. Zheng. Structured annotations to support collaborative writing workflow. Master's thesis, University of British Columbia, 2005.

[43] Q. Zheng, K. S. Booth, and J. McGrenere. Co-authoring with structured annotations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 131–140. ACM Press, 2006.

A.2 Study Document

Challenges of CSCW

The term CSCW was originally coined by Irene Greif at a workshop in 1984 attended by a small group interested in using technology to support people at their work, which led to the first CSCW conference in 1986 [7]. CSCW focuses on the study of tools and applications to support groups and their social, psychological, and organizational impacts. The term CSCW stands for Computer Supported Cooperative Work. Tools and applications in the CSCW context mainly support the coordination and communication for small groups or organizations such as scheduling meetings, distributed decision making, locating colleagues, collaborative work, etc. Since then, CSCW has evolved as a field with many influences from various disciplines such as: Computer Science, Sociology, Psychology, Ethnography, Anthropology, Organizational Studies and Communications. The field has accumulated a set of empirically based interdisciplinary studies, and many new interesting applications to support groups and organizations. The design and development of the groupware tools involve challenges regarding not only technical issues, but also many other factors.

Applications designed only with a technology-centered perspective and without much consideration for users can result in neither usable nor useful applications, no matter how innovative or constructive the technology is. The human computer interaction researchers address and solve such a problem by proposing user-centered interface design guidelines or heuristics for usable and efficient applications [13]. However, such individual-centered approaches fail to address issues involving multi-user applications. The CSCW field has proposed organization-centered approaches, by broadened design perspectives to address social and organizational contexts of technology use in groups and organizations. Examination and consideration of group's nature, work practice, interaction, and incentives become essential in designing and developing groupware applications.

Designing and deploying a successful groupware application involves many challenges. In addition to be usable and efficient for the common goal of the group, the applications are expected to benefit the individuals of the group. The developers have to foresee and consider the potential reciprocal and co-evolutionary relationship between the technology and social context of the group: the use context is effected by the constraints

and affordances of the technology offered by the application and the technology also co-adapts to the environment in important ways [14]. In order to develop a usable and efficient groupware application, developers need to face and cope with some of the major fundamental challenges. A critical mass of use of the new technology or application is often required to gain the total benefit of technology [6].

As discussed earlier, most groupware require the critical mass of use of the application to succeed. However, most groups constitute heterogeneous mixtures of members with different backgrounds and preferences. Therefore, groupware should be designed so that it is appealing to users with various background and knowledge. One possible solution is the use of adaptable user interfaces so that every group member can customize the application to their needs and preferences [11]. This is a problem and we need to deal with it wisely. For single-user application, a word processor liked by one in five prospective customers could be a huge success [5]. However, for multi-user application, an application liked by only one member out of five in the group will be a failure since the decisions have to be made at the group level. Thus the groupware should be appealing to all the group members.

Reduce Friction

More often than not, organizations and groups are structured and responsibilities are divided so that the overhead for communications will be minimized. However, as the size of the group grows bigger the more overhead requirements for communication and coordination it demands. Some groupware applications are designed mainly to provide such communication and coordination needs among the group members. One of the examples is a groupware with features that support collaborative co-authoring, version tracking, and distribution of drafts for collaborative tasks. However, if a collaborative groupware and users' main application are not integrated together, the users will need to leave their core applications and launch such collaborative tools or visit collaborative platform to get the latest version of the shared artifact. Consequently, the "friction" or unnecessary delays for users' work [8]. Groupware should be designed to minimize or eliminate the extra work, and to give users maximum seamless working experience. Hence, groupware features will fare better if they are integrated with applications that users use for their primary tasks [5].

Privacy, Social and Political Issues

The technology and functionalities offered by some groupware can lead to an activity that causes conflicts, effects personal privacy of individuals, interferes with social dynamics, or threatens existing political structure of the group. In everyday situations, we play it by ear according to social protocols, temporal situations, tactical agendas and awareness of the personalities. However, groupware applications often require making such protocols and information explicit, leaving no room for tactical procedure or improvisation. For example: a groupware calendar system and automatic scheduling system that is open for the group members to view and that allows priority-based scheduling will result in failure because users would be reluctant to publicly acknowledge that some of their meetings with other group members are low priority [5]. Therefore, during the design process, groupware developers should consider those social and political aspects of groups and consequences that the applications can cause and how to prevent them.

Failure of Intuitive Decision-Making

Often, the responsible manager who comes to a decision about resources for an application development project relies heavily on intuition [4]. We can use intuition as a guide for decisions about single user applications because it might not be very difficult for the decision-maker to imagine ourselves with the target users. However, for groupware, it needs more than intuition; in fact, education and vigilance about the nature of groups and different interests and incentives of each member [4]. Additionally, the manager might even come across issues involving conflicts of interest or bias towards a certain subset of the user population because groupware applications are intended for users with different job titles, heterogeneous background and various interests.

Evaluation

Evaluation for CSCW applications requires a very different approach and methodology than single user applications. Obstacles to achieve generalizability, precision and realism exist due to various potential impacts such as social, physical, political, temporal, motivational and economic involved in a group, central to the group's performance and success [4]. Most evaluations and studies involve a very small group with constant membership, performing simple tasks arbitrarily assigned to them by the experimenter under "context-stripped" conditions [9]. Just as Gould's "How to Design Usable

Systems" [3] and other usability design guidelines [1], systematic and comprehensive design models and guidelines to include group-centered perceptive will be helpful for developers. Moreover, exploration and identification of appropriate research methodologies for better evaluations of groupware will result in more successful and usable groupware applications.

CASE STUDY: *Microsoft SCHEDULE+*

The collaborative meeting scheduler application SCHEDULE+ is widely used by employees at the Microsoft Corporation, as Grudin and Palen reported from their three month long study as a participant observer in a Microsoft development division [6]. It is successful now, but it was a failure when first introduced in the early 1980s [4]. The recent success of the application mainly contributes to the social context of the group adapted by the application, and better design and group-centered perspectives from the developers.

Overall, SCHEDULE+ has evolved as a successful groupware application over ten years since it was first introduced. The users at Microsoft seem to have adapted well to the application and its features offered. Moreover, its success could also be due to the fact that developers have learned and become aware of the problems and issues around such applications and overcome most of them.

Is CSCW Ready for Theory?

Since CSCW has emerged, some theories have imposed influence on CSCW. An example of such theories is Structuration Theory, which states that the social context is effected by the new technology introduced and that the technology implementation is in turn influenced by users [9]. Moreover, other theories have also been proposed to solve the problems and issues surrounding groupware [9]. The recently debating question is whether CSCW as a field is ready for theory: whether the technology and problems are mature enough to accommodate theory or whether theory is needed to advance the field [9].

Where is CSCW going?

CSCW has first emerged as a field including people with different background and yet common interest to utilize technology to support group work. As Grudin noted for CSCW,

> *"People come from different places, and it is useful to know where each is from and why they have come. Each visitor can see what the others have to offer and can decide what is worth taking home. [...] There is no assumption that everyone speaks the same language, only that they will try to work out some means of communicating. [...] When understood and respected, the differences form the core of richer, shared understandings."*

Consequently, CSCW has grown as a unique multidisciplinary field. Researchers have contributed multiple perspectives and insights for the growth of CSCW field and bigger success of CSCW applications.

A.3 Interview Questions

1. How long have you been tutoring?
2. How many documents have you reviewed as a tutor?
3. How many hours did you dedicate to tutoring (reviewing documents) on a weekly basis?
4. What types of documents were usually asked to review? (e.g., essays, term papers, thesis, publishable papers)
5. On average, how long were documents? (single spaced, with standard font size 9-12)
6. How do you usually review the documents?
    a. Print out (Rarely ----- Frequently)
        i. Why?
        ii. Does this make the reviewing process easier or more efficient than other options?
        iii. Positive aspects?
        iv. Negative aspects?
    b. Directly edit using word processor (Rarely ----- Frequently)
        i. Why?
        ii. Does this make the reviewing process easier or more efficient than other options?
        iii. Positive aspects?
        iv. Negative aspects?
    c. Use annotation function (Rarely ----- Frequently)
        i. Why?
        ii. Does this make the reviewing process easier or more efficient than other options?
        iii. Positive aspects?
        iv. Negative aspects?
    d. Write email message (Rarely ----- Frequently)
        i. Why?
        ii. Does this make the reviewing process easier or more efficient than other options?
        iii. Positive aspects?
        iv. Negative aspects?
    e. Online communication groupware (Rarely ----- Frequently)
        i. Why?
        ii. Does this make the reviewing process easier or more efficient than other options?
        iii. Positive aspects?
        iv. Negative aspects?
    f. Other ?
        i. Why?
        ii. Does this make the reviewing process easier or more efficient than other options?
        iii. Positive aspects?
        iv. Negative aspects?
7. On average, how long did you take to review a document (per page)?

8. On average, how many passes did you take at a document in the process of reviewing it?
9. Is there any particular strategy that you used while reviewing a document? (e.g., grammar first, style issues second, etc.)
10. What aspects of documents did you mainly comment on? Did students have to request you to make those comments? Upon request only or "voluntary"?
    a. Grammar
    b. Style
    c. Structure
    d. Clarity issues
    e. Resolve conflicts
    f. Other
11. On average, after reviewing, how many comments and edits appeared per page?
12. After reviewing the document, what else did you communicate to the writer(s)? (e.g, high level comments, prioritized problems, etc.)
13. How did you communicate to the writer(s)?
    a. Email
        i. Why?
        ii. Does this make the reviewing process easier or more efficient than other options?
        iii. Positive aspects?
        iv. Negative aspects?
    b. Face to face
        i. Why?
        ii. Does this make the reviewing process easier or more efficient than other options?
        iii. Positive aspects?
        iv. Negative aspects?
    c. Document drop off
        i. Why?
        ii. Does this make the reviewing process easier or more efficient than other options?
        iii. Positive aspects?
        iv. Negative aspects?
    d. Phone
        i. Why?
        ii. Does this make the reviewing process easier or more efficient than other options?
        iii. Positive aspects?
        iv. Negative aspects?
    e. Others
        i. Why?
        ii. Does this make the reviewing process easier or more efficient than other options?
        iii. Positive aspects?
        iv. Negative aspects?

14. How long did it take to communicate with the writers? (longer than reviewing time, etc.)
15. When were the documents requested to be reviewed?
    a. Early draft
    b. Towards the middle of draft
    c. Almost ready draft
16. Were there due dates for the documents that you were requested to review?
17. How many iterations did you review a document? Any follow up?
18. Are there any (technological as well as intellectual) restrictions that were imposed by the program on how you review the documents?

B.2 Demographic Questionnaire

Using Structured Annotations in Collaborative Writing
Study Questionnaire Form

Instructions

Please try to respond to all of the items listed below. For those items that are not applicable, specify **N/A**.

Past Computer and Writing Experience *(To be completed before the study)*

1. Which word processor do you currently use for writing documents (e.g. essays, reports, letters, conference papers, journal articles, etc.)?

2. How often do you use the word processor?

☐ Once a month ☐ Once a week ☐ Every 2-3 days ☐ Every day

3. How confident do you feel about using the word processor?

(Not at all confident) 1 2 3 4 5 6 7 8 9 10 (Very confident)

4. Do you use the annotation functions in the word processor? (e.g. Track Changes and Commenting functions in Microsoft Word)

☐ Yes.

☐ No, please specify why:_____

5. Have you previously written or reviewed documents with other people?

☐ None ☐ less than 5 times ☐ between 5~10 times ☐ more than 10 times

Word processor used in collaborative writing: _____

6. Please describe briefly the nature of the documents that you have co-authored (e.g., project reports for a course or job, conference papers, journal papers).

(Continue on the next page)

7. How do you and your co-authors review a collaborative document? (Check all the items that apply.)

Reviewing Method	Frequency of use (Mark an 'X' along the scale)
☐ Print out the document, mark on the document using a pen, and then hand back the marked document to co-author(s).	(Rarely) (Frequently)
☐ Directly edit on the document using a word processor.	(Rarely) (Frequently)
☐ Use the annotation function in the word processor to edit the document and add comments.	(Rarely) (Frequently)
☐ Write an email message that includes suggested changes and comments about the document to other co-authors.	(Rarely) (Frequently)
☐ Use online communication groupware (e.g., Yahoo! Groups) to discuss about the changes to the document.	(Rarely) (Frequently)
☐ Other. Please specify:_____	(Rarely) (Frequently)

Do not turn over the page. Please notify the experimenter when you finish answering the above questions.

B.3 Training Task

Scenario:

The followings are the items you need to buy.

1. Beef
2. Broccoli
3. Grapes
4. Milk
5. Chicken
6. Cheese
7. Banana
8. Yogurt
9. Eggs

Your roommate Nancy and Amber are going for grocery shopping and have kindly offered to help you pick up the items you need. So, your task is to organize the items so that they can efficiently do the shopping.

B.4 Instructions to Participants

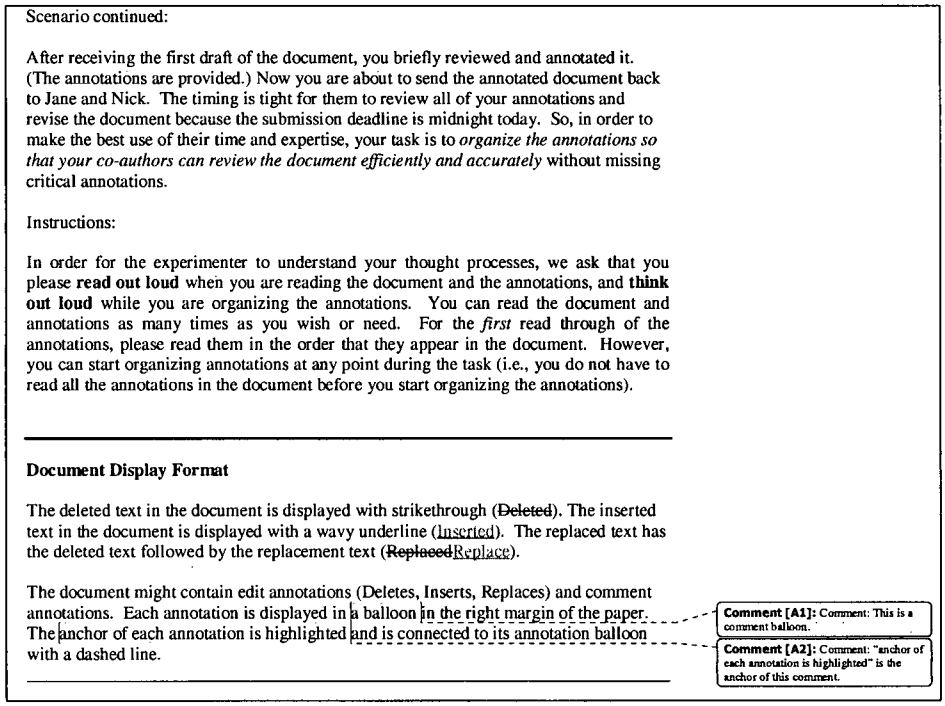
Scenario:

Imagine you are Sam, and are supervising Jane and Nick on a project about "music effects on memorization". You are not directly involved in the project, but rather you provide directions to Jane and Nick and supervise the project at a higher level. All three of you are currently collaborating on a document titled "The Effects of Music Type and Volume on Memorization". Thus, Jane and Nick are your co-authors, who are equally responsible for editing the document. Their roles in the project and in writing this document are as followings:

Jane's role: Domain expert in effects of music on cognitive processes
Nick's role: Expert in experimental design and analysis

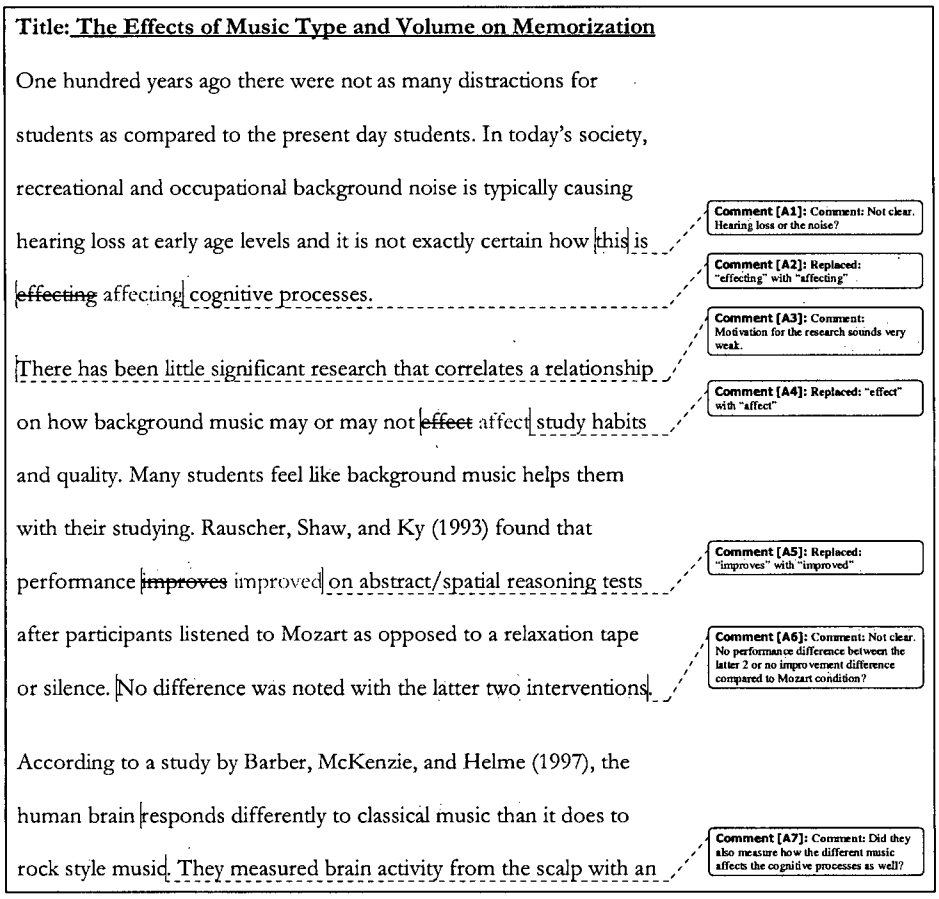
You have received their first draft of the document. Now, you have 5 minutes to skim through the document to familiarize yourself with what they have written. Do not annotate or comment on the document at this point. Annotations and comments will be provided to you after you have read the document.

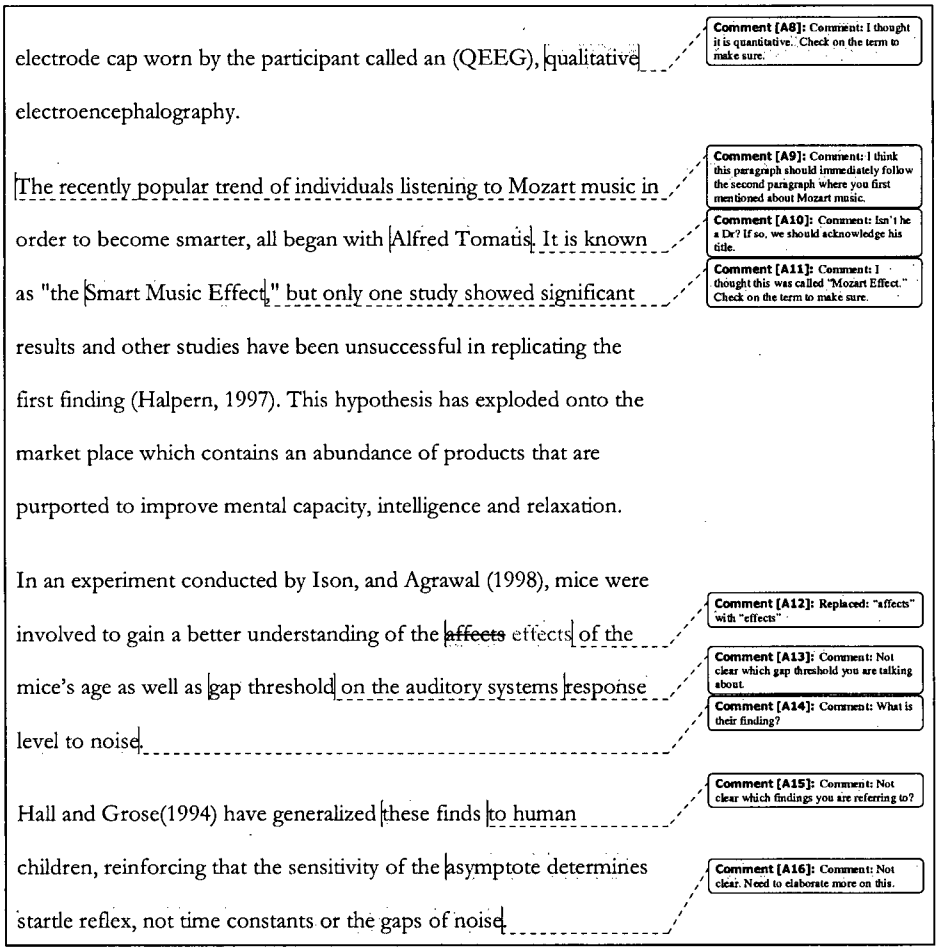
Please notify the experimenter when you finish reading the above scenario.

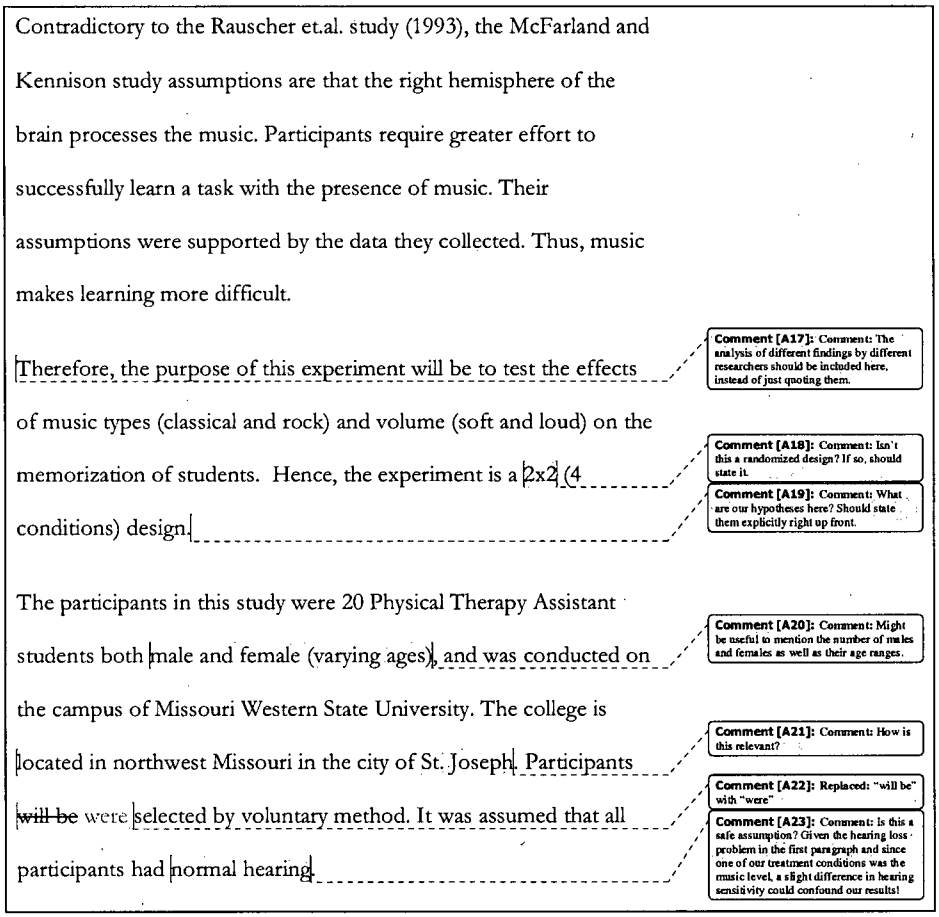


B.5 Study Document with Annotations

The document was selected from an online archive of manuscripts from the National Undergraduate Research Clearinghouse of Missouri Western State University: http://clearinghouse.missouriwestern.edu/


Title: The Effects of Music Type and Volume on Memorization
One hundred years ago there were not as many distractions for students as compared to the present day students. In today's society, recreational and occupational background noise is typically causing hearing loss at early age levels and it is not exactly certain how this is ~~effecting~~ affecting cognitive processes.
There has been little significant research that correlates a relationship on how background music may or may not ~~effect~~ affect study habits and quality. Many students feel like background music helps them with their studying. Rauscher, Shaw, and Ky (1993) found that performance ~~improves~~ improved on abstract/spatial reasoning tests after participants listened to Mozart as opposed to a relaxation tape or silence. No difference was noted with the latter two interventions.
According to a study by Barber, McKenzie, and Helme (1997), the human brain responds differently to classical music than it does to rock style music. They measured brain activity from the scalp with an
Comment [A1]: Comment: Not clear. Hearing loss or the noise?
Comment [A2]: Replaced: "effecting" with "affecting"
Comment [A3]: Comment: Motivation for the research sounds very weak.
Comment [A4]: Replaced: "effect" with "affect"
Comment [A5]: Replaced: "improves" with "improved"
Comment [A6]: Comment: Not clear. No performance difference between the latter 2 or no improvement difference compared to Mozart condition?
Comment [A7]: Comment: Did they also measure how the different music affects the cognitive processes as well?


electrode cap worn by the participant called an (QEEG), qualitative electroencephalography.
Comment [A8]: Comment: I thought it is quantitative. Check on the term to make sure.
The recently popular trend of individuals listening to Mozart music in order to become smarter, all began with Alfred Tomatis. It is known as "the Smart Music Effect," but only one study showed significant results and other studies have been unsuccessful in replicating the first finding (Halpern, 1997). This hypothesis has exploded onto the market place which contains an abundance of products that are purported to improve mental capacity, intelligence and relaxation.
Comment [A9]: Comment: I think this paragraph should immediately follow the second paragraph where you first mentioned about Mozart music.
Comment [A10]: Comment: Isn't he a Dr? If so, we should acknowledge his title.
Comment [A11]: Comment: I thought this was called "Mozart Effect." Check on the term to make sure.
In an experiment conducted by Ison, and Agrawal (1998), mice were involved to gain a better understanding of the affects effects of the mice's age as well as gap threshold on the auditory systems response level to noise.
Comment [A12]: Replaced: "affects" with "effects"
Comment [A13]: Comment: Not clear which gap threshold you are talking about.
Comment [A14]: Comment: What is their finding?
Hall and Grose(1994) have generalized these finds to human children, reinforcing that the sensitivity of the asymptote determines startle reflex, not time constants or the gaps of noise.
Comment [A15]: Comment: Not clear which findings you are referring to?
Comment [A16]: Comment: Not clear. Need to elaborate more on this.


Contradictory to the Rauscher et.al. study (1993), the McFarland and Kennison study assumptions are that the right hemisphere of the brain processes the music. Participants require greater effort to successfully learn a task with the presence of music. Their assumptions were supported by the data they collected. Thus, music makes learning more difficult.
Therefore, the purpose of this experiment will be to test the effects of music types (classical and rock) and volume (soft and loud) on the memorization of students. Hence, the experiment is a 2x2 (4 conditions) design.
The participants in this study were 20 Physical Therapy Assistant students both male and female (varying ages), and was conducted on the campus of Missouri Western State University. The college is located in northwest Missouri in the city of St. Joseph. Participants will be were selected by voluntary method. It was assumed that all participants had normal hearing.
Comment [A17]: Comment: The analysis of different findings by different researchers should be included here, instead of just quoting them.
Comment [A18]: Comment: Isn't this a randomized design? If so, should state it.
Comment [A19]: Comment: What are our hypotheses here? Should state them explicitly right up front.
Comment [A20]: Comment: Might be useful to mention the number of males and females as well as their age ranges.
Comment [A21]: Comment: How is this relevant?
Comment [A22]: Replaced: "will be" with "were"
Comment [A23]: Comment: Is this a safe assumption? Given the hearing loss problem in the first paragraph and since one of our treatment conditions was the music level, a slight difference in hearing sensitivity could confound our results!

Items included in this experiment were a Realistic soundlevel meter 33-2080, Aiwa brand compact disk stereo player, 2 compact disks, and a memory test. The memory test consisted of two different word lists of 25 capital cities in the United States. The classical music and the heavy metal music were played.

Comment [A24]: Comment: Only two different lists? We have 4 conditions. Unless the presentation of the list is counterbalanced, this could be a major confounding factor in our results!!

Comment [A25]: Comment: State which song we used for classical in the experiment.

Comment [A26]: Comment: State which song we used for heavy metal music in the experiment.

Testing occurred in 4, twenty-minute sessions. Participants entered the test room with one of the possible four interventions in progress. Heavy metal style music with volume levels soft or loud or classical style music with volume levels set at soft or loud. Soft music ~~is~~ was played at 55 dB and loud music was played at 70 dB both musical sources were 10 feet away from the participants. A memory test was present, face down at the participant's seat. The researcher instructed the participants that on the cue they were to turn over the word list in front of them and given two minutes to review it. On the second cue they would return list face down and were then given three minutes to recall as many words as possible.

Comment [A27]: Comment: Why in progress? When are the instructions given to the subjects?

Comment [A28]: Replaced: "is" with "was"

Comment [A29]: Comment: Try avoiding short forms – state the actual units.

Comment [A30]: Comment: Was the music still playing during the recall as well?

The numbers of correct responses made by the subjects were recorded with the two types of interventions that were present at test time.

A 2X2 factorial ANOVA were calculated comparing music type and volume to memory test scores. The ANOVA test was calculated in order to compare music volume levels and music type on correct test responses. The main effect for volume level on correct responses was not significant ($F(1,36)= 1.26, p= .27$). The main effect for music type on the number of correct responses was not significant ($F(1,36)=.686, p = .413$).The volume by music type interaction was not significant ($F(1,36) =.116$ $p=.735$).Thus it appears that neither music type or volume has any effect on the number of correct responses on a memory test.

Initially, it was expected that music type and volume would have a strong relationship to students' performance on the memory test. Statistical analysis showed a weak relationship between music type and test performance. There was a slight decrease in the test scores of students exposed to high volume music levels. They had an increase

Comment [A31]: Comment: Isn't it common in the literature to record both correct and incorrect responses for the memory recall tests?

Comment [A32]: Comment: Two or Four?

Comment [A33]: Comment: This is within subject design, right? Need to mention within subject or between subject explicitly here.

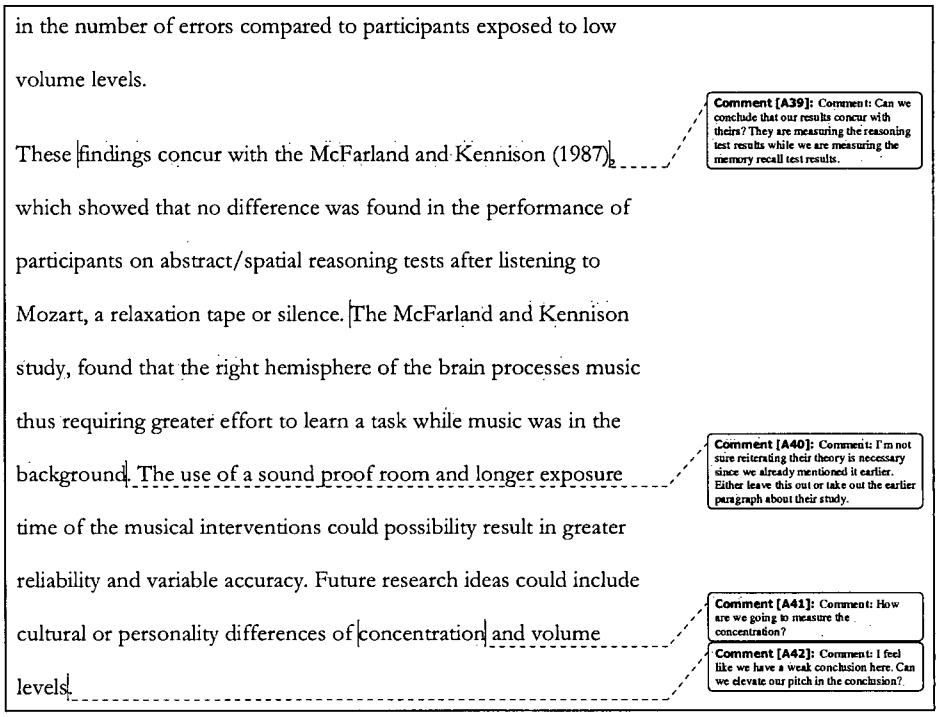
Comment [A34]: Comment: I think it'll be useful to give the data about the average number of correct responses and errors made by the subjects.

Comment [A35]: Comment: Should report standard error.

Comment [A36]: Comment: Should report standard error.

Comment [A37]: Comment: Should report standard error.

Comment [A38]: Comment: What theory is this hypothesis based on? We should explain why we expected that way.


in the number of errors compared to participants exposed to low volume levels.
These findings concur with the McFarland and Kennison (1987), which showed that no difference was found in the performance of participants on abstract/spatial reasoning tests after listening to Mozart, a relaxation tape or silence. The McFarland and Kennison study, found that the right hemisphere of the brain processes music thus requiring greater effort to learn a task while music was in the background. The use of a sound proof room and longer exposure time of the musical interventions could possibility result in greater reliability and variable accuracy. Future research ideas could include cultural or personality differences of concentration and volume levels.
Comment [A39]: Comment: Can we conclude that our results concur with theirs? They are measuring the reasoning test results while we are measuring the memory recall test results.
Comment [A40]: Comment: I'm not sure reiterating their theory is necessary since we already mentioned it earlier. Either leave this out or take out the earlier paragraph about their study.
Comment [A41]: Comment: How are we going to measure the concentration?
Comment [A42]: Comment: I feel like we have a weak conclusion here. Can we elevate our pitch in the conclusion?

C.2 Questionnaires

C.2.1 Demographic Questionnaire

UBC	***Collaborative Writing with Annotations*** ***Study Questionnaire Form***

Instructions

Please try to respond to all of the items listed below. For those items that are not applicable, specify **N/A**.

Part 1: Past Computer, Reviewing and Co-authoring Experience *(To be completed before the study)*

Computer Experience

1. Which word processor do you currently use for writing/reviewing documents (e.g. essays, reports, letters, conference papers, journal articles, etc.)?

2. How often do you use the word processor?

☐ Once a month ☐ Once a week ☐ Every 2-3 days ☐ Every day

3. How confident do you feel about using the word processor?

(Not at all confident) 1 2 3 4 5 6 7 (Very confident)

4. Do you use the annotation functions in the word processor? (e.g., Track Changes and Commenting functions in Microsoft Word)

☐ Yes.

☐ No, please specify why:_____

Reviewing Experience

5. How many times have you previously reviewed documents *written by other people*?

☐ None ☐ less than 5 times ☐ between 5~10 times ☐ more than 10 times

Word processor used in reviewing: _____

6. Please describe briefly the nature of the documents that you have reviewed (e.g., project reports for a course or job, conference papers, journal papers).

7. How do you review documents? (Check all the items that apply.)

Reviewing Method	Frequency of use (Mark an 'X' along the scale)
☐ Print out the document, mark on the document using a pen, and then hand back the marked document to author(s).	(Rarely) (Frequently)
☐ Directly edit on the document using a word processor.	(Rarely) (Frequently)
☐ Use the annotation functions in the word processor to edit the document and add comments.	(Rarely) (Frequently)
☐ Write an email message that includes suggested changes and comments about specific parts of the document to author(s).	(Rarely) (Frequently)
☐ Other. Please specify:_____	(Rarely) (Frequently)

8. How do you communicate/discuss the annotations and reviewing results with the author(s)? (check all the items that apply)

Communication Method	Frequency of use (Mark an 'X' along the scale)
☐ Face to face meetings.	(Rarely) (Frequently)
☐ Email.	(Rarely) (Frequently)
☐ Instant Messaging.	(Rarely) (Frequently)
☐ Phone.	(Rarely) (Frequently)
☐ Use online communication groupware (e.g., Yahoo! Groups) to discuss about the changes to the document.	(Rarely) (Frequently)
☐ Other. Please specify:_____	(Rarely) (Frequently)

Co-authoring Experience

9. Have you previously co-authored (written and reviewed) documents *with other people*?

☐ None ☐ less than 5 times ☐ between 5~10 times ☐ more than 10 times

Word processor used in co-authoring: _____

10. Please describe briefly the nature of the documents that you have co-authored (e.g., project reports for a course or job, conference papers, journal papers).

11. How do you discuss about collaborative documents? (Check all the items that apply.)

Reviewing Method	Frequency of use (Mark an 'X' along the scale)
☐ Print out the document, mark on the document using a pen, and then hand back the marked document to co-author(s).	(Rarely) (Frequently)
☐ Directly edit on the document using a word processor.	(Rarely) (Frequently)
☐ Use the annotation functions in the word processor to edit the document and add comments.	(Rarely) (Frequently)
☐ Write an email message that includes suggested changes and comments about specific parts of the document to co-author(s).	(Rarely) (Frequently)
☐ Use online communication groupware (e.g., Yahoo! Groups) to discuss about the changes to the document.	(Rarely) (Frequently)
☐ Other. Please specify:_____	(Rarely) (Frequently)

Do not turn over the page. Please notify the experimenter when you finish answering the above questions.

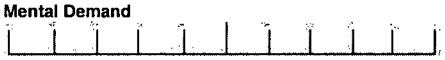
C.2.2 System-specific Questionnaire: Bundle System

Experience: *(To be completed after completing the task using the bundle system)*

With respect to your **experience performing the given task with the bundle system**, please answer the following questions by marking an 'X' along the scale below the corresponding question.

1. How much mental and perceptual activity was required to perform the task (e.g., thinking, deciding, remembering, searching, etc.)?

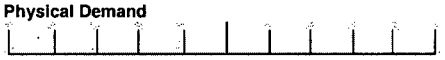
   Mental Demand

   (Low) (High)

2. How much physical activity was required (e.g., scrolling, clicking, typing, etc.)?

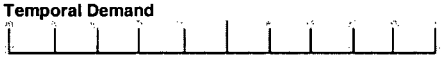
   Physical Demand

   (Low) (High)

3. How much time pressure did you feel due to the rate or pace at which the tasks needed to be accomplished?

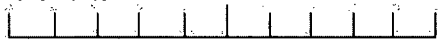
   Temporal Demand

   (Low) (High)

4. How successful do you think you were in accomplishing the goals of the task set by the experimenter?

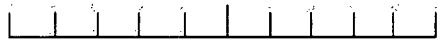
   Performance

   (Low) (High)

5. How hard did you have to work (mentally and physically) to accomplish your level of performance?

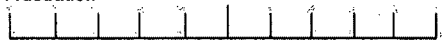
   Effort

   (Low) (High)

6. How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

   Frustration

   (Low) (High)

Importance of Different Workload Categories

Please select the member of each pair that had the more significant effect on the overall workload for the task that you just performed using the filer system and the email. Descriptions of the categories can be found at the bottom of the page.

Physical Demand	☐ or ☐	Mental Demand
Temporal Demand	☐ or ☐	Mental Demand
Performance	☐ or ☐	Mental Demand
Effort	☐ or ☐	Mental Demand
Frustration Level	☐ or ☐	Mental Demand
Temporal Demand	☐ or ☐	Physical Demand
Performance	☐ or ☐	Physical Demand
Effort	☐ or ☐	Physical Demand
Frustration Level	☐ or ☐	Physical Demand
Temporal Demand	☐ or ☐	Performance
Temporal Demand	☐ or ☐	Effort
Temporal Demand	☐ or ☐	Frustration Level
Performance	☐ or ☐	Effort
Performance	☐ or ☐	Frustration Level
Frustration Level	☐ or ☐	Effort

Mental Demand: How much mental and perceptual activity was required to perform the task (e.g., thinking, deciding, remembering, searching, etc.)?

Physical Demand: How much physical activity was required (e.g., scrolling, clicking, typing, etc.)?

Temporal Demand: How much time pressure did you feel due to the rate or pace at which the tasks needed to be accomplished?

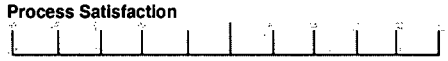
Performance: How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)?

Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?

Frustration: How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

7. How satisfied are you with the process taken to accomplish the task?

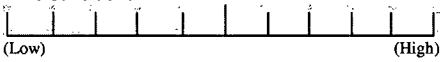
Process Satisfaction



(Low) (High)

8. How satisfied are you with the time taken to accomplish the task?

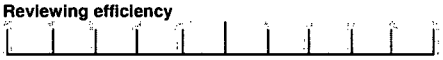
Time Satisfaction



(Low) (High)

9. To what extent do you think your annotations and bundles will facilitate the efficient reviewing by your co-authors?

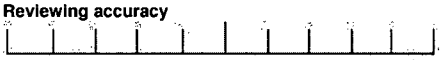
Reviewing efficiency



(Low) (High)

10. To what extent do you think your annotations and bundles will facilitate the accurate reviewing by your co-authors?

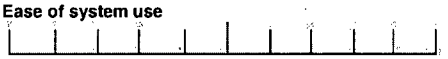
Reviewing accuracy



(Low) (High)

11. How satisfied are you with the ease of using this writing system?

Ease of system use



(Low) (High)

Questions:

1. What was the most difficult task for you to complete using the system?

2. What particular aspect(s) of *bundles* did you **like**?

3. What particular aspect(s) of *bundles* did you **dislike?**

4. What particular aspect(s) of this *collaborative writing system* did you **like**?

5. What particular aspect(s) of this *collaborative writing system* did you **dislike**?

C.2.3 System-specific Questionnaire: Simple System

Experience: *(To be completed after completing the task using the filter system)*

With respect to your **experience performing the given task with the filter system and email**, please answer the following questions by marking an 'X' along the scale below the corresponding question.

1. How much mental and perceptual activity was required to perform the task (e.g., thinking, deciding, remembering, searching, etc.)?

   Mental Demand

   (Low) (High)

2. How much physical activity was required (e.g., scrolling, clicking, typing, etc.)?

   Physical Demand

   (Low) (High)

3. How much time pressure did you feel due to the rate or pace at which the tasks needed to be accomplished?

   Temporal Demand

   (Low) (High)

4. How successful do you think you were in accomplishing the goals of the task set by the experimenter?

   Performance

   (Low) (High)

5. How hard did you have to work (mentally and physically) to accomplish your level of performance?

   Effort

   (Low) (High)

6. How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

   Frustration

   (Low) (High)

Importance of Different Workload Categories

Please select the member of each pair that had the more significant effect on the overall workload for the task that you just performed using the filer system and the email. Descriptions of the categories can be found at the bottom of the page.

Physical Demand	☐ or ☐	Mental Demand
Temporal Demand	☐ or ☐	Mental Demand
Performance	☐ or ☐	Mental Demand
Effort	☐ or ☐	Mental Demand
Frustration Level	☐ or ☐	Mental Demand
Temporal Demand	☐ or ☐	Physical Demand
Performance	☐ or ☐	Physical Demand
Effort	☐ or ☐	Physical Demand
Frustration Level	☐ or ☐	Physical Demand
Temporal Demand	☐ or ☐	Performance
Temporal Demand	☐ or ☐	Effort
Temporal Demand	☐ or ☐	Frustration Level
Performance	☐ or ☐	Effort
Performance	☐ or ☐	Frustration Level
Frustration Level	☐ or ☐	Effort

Mental Demand: How much mental and perceptual activity was required to perform the task (e.g., thinking, deciding, remembering, searching, etc.)?

Physical Demand: How much physical activity was required (e.g., scrolling, clicking, typing, etc.)?

Temporal Demand: How much time pressure did you feel due to the rate or pace at which the tasks needed to be accomplished?

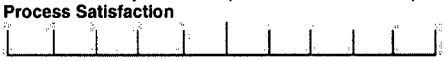
Performance: How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)?

Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?

Frustration: How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

7. How satisfied are you with the process taken to accomplish the task?

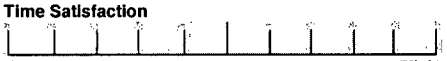
Process Satisfaction



(Low) (High)

8. How satisfied are you with the time taken to accomplish the task?

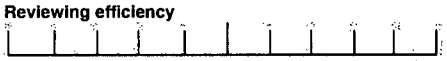
Time Satisfaction



(Low) (High)

9. To what extent do you think your annotations and email message will facilitate the efficient reviewing by your co-authors?

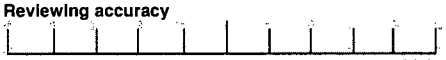
Reviewing efficiency



(Low) (High)

10. To what extent do you think your annotations and email message will facilitate the accurate reviewing by your co-authors?

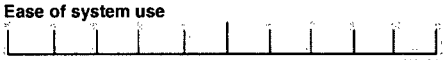
Reviewing accuracy



(Low) (High)

11. How satisfied are you with the ease of using this writing system?

Ease of system use



(Low) (High)

Questions:

1. What was the most difficult task for you to complete using the system?

2. What particular aspect(s) of this *collaborative writing system* did you **like**?

3. What particular aspect(s) of this *collaborative writing system* did you **dislike**?

C.2.4 Comparative Questionnaire

Comparative Experience: *(To be completed after completing both tasks)*

Please answer the following questions by marking
- an 'X' along the scale with respect to your experience performing the task with the bundle system,
- an 'O' along the scale with respect to your experience performing the task with the filter system and email.

We are interested in your experiences with each system relative to the other. Hence, you do not have to worry about your answers in this questionnaire identical to your answers in the previous questionnaires.

1. How much mental and perceptual activity was required to perform the task (e.g., thinking, deciding, remembering, searching, etc.)?

   Mental Demand

   (Low) (High)

2. How much physical activity was required (e.g., scrolling, clicking, typing, etc.)?

   Physical Demand

   (Low) (High)

3. How much time pressure did you feel due to the rate or pace at which the tasks needed to be accomplished?

   Temporal Demand

   (Low) (High)

4. How successful do you think you were in accomplishing the goals of the task set by the experimenter?

   Performance

   (Low) (High)

5. How hard did you have to work (mentally and physically) to accomplish your level of performance?

   Effort

   (Low) (High)

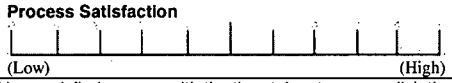
6. How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

   Frustration

   (Low) (High)

7. How satisfied are you with the process taken to accomplish the task?

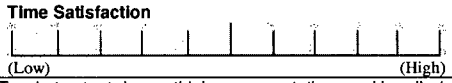
Process Satisfaction



(Low) (High)

8. How satisfied are you with the time taken to accomplish the task?

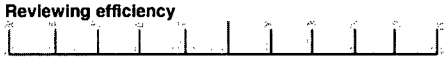
Time Satisfaction



(Low) (High)

9. To what extent do you think your annotations and bundle / email message will facilitate the efficient reviewing by your co-authors?

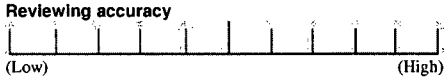
Reviewing efficiency



(Low) (High)

10. To what extent do you think your annotations and bundle / email message will facilitate the accurate reviewing by your co-authors?

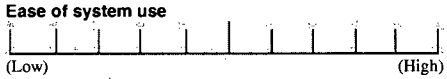
Reviewing accuracy



(Low) (High)

11. How satisfied are you with the ease of using this writing system?

Ease of system use



(Low) (High)

C.3 Study Documents

C.3.1 Training Document

Parenting and Honesty

Some parenting experts urge parents to tell their children the truth about their past experiment with drugs and sex, even when they don't want their children to do what they did. Others think parents should censor what they tell their children. This paper will demonstrate that it is better for parents to censor what they tell their children until the children are old enough to understand.

One of the primary duties of parenting is to impart the parents' moral views to their children. Moral views are developed through a lifetime. This means that the activities that a parent engaged in, especially with drugs and sex, long before he or she was a parent, formed his or her moral views. This does not mean that they are stuck with these views, but that they helped to shape the parents' views. For instance, a woman who experienced with drugs as a teenager and regrets it will probably have developed the moral view that teenagers should not take drugs, based on (not despite) the fact that she did it herself. Children and young teens just don't get it, so it would be harmful for the parent to have to hide her drug use while telling her children not to try drugs.

Children have such a black-and-white view of the world that they cannot understand that a person could do something and then regret it later. Confessing all of one's young indiscretions would make a parent's discussions with their children about drugs and sex counterproductive. In fact, many parents might choose not to discuss these subjects at all, rather than risk confusing their children or revealing facts their children are too young to know.

Opponents of this view might argue that once a child learns the truth about his or her parents' activities, he or she will defeat the parent and will no longer see the parents' views as credible. However, this is not a logical assumption, since by the time the child finds out the truth, he or she will be old enough to understand why the parent lied. A 22-year-old who finds out that her mother smoked pot in college will understand why the mother did not tell her that when she was 12. An 18-year-old who discovers that his

father has premarital sex will most likely feel very differently about that than he would have when he was 11.

To conclude, parents need to adjust what they told their children to fit the child's age and developmental stage. When the children are young, they should be less concerned with telling the absolute truth about their own experiences than they should be with making sure their children know what their own moral views are. Later, once the children are old enough, the parents could choose to reveal the truth of their own experiments, but at this point the children will be past the danger point.

C.3.2 Task Documents

Document 1: DocH

Extending high school to five years

1. Introduction

"Senior itch" is an incessant irritation that high school students all crave to scratch. The cure? Graduation. As students progress through their high school years growing in wisdom and maturity, they are desperate for freedom. Yet what they desire most is not always what is best for them. Although most students don't wish to admit it, extending the high school career to five years would make an important and beneficial affect on their future. The current four years of high school education do not provide sufficient time for motivated students to accomplish all of their goals before college. Merely getting accepted by a selective college or university requires much pre-planned effort, and students who are preoccupied with their studies and other activities simply do not have enough time.

2. Interest

High schools are always adding new and interesting courses to the curricula which they offer. However, many college-bound students do not have sufficient time to explore these different classes. High school students thus are forced to choose classes that fit into their limited schedules. Occasionally classes that perceive to be easier are chosen, merely to make sure that their transcripts stand out in the admissions pool. Fifteen years of age seems too late to start making choices based on "career", rather than choosing classes that they might find interesting and fun. An extra year of high school would provide students with opportunities to explore different subjects and take classes that they interest. Consequently, students would enjoy their education and develop more stakes in school.

3. Individuality

Colleges thoroughly consider the ways in which an applicant has spent his or her four years of high school. Students who take on leadership roles, display dictation to an organization, and show other evidence of being well-rounded appeal most to elite educational institutions. Although some high school students may desire to take on leadership positions numerous extra-curricular organizations, they are often faced with school regulations that limit the number of offices they may hold. Even when their school does not place restrictions on such involvement, students eventually reach the limits of what can be done in a 24-hour day. All too often, students will find that they cannot participate in as many extra-curricular activities as they would like, due to the fact that they just do not have enough time. An extra year of high school would enable such students to become involved in more than one activity, providing them with sufficient time to become fully engaged in each to a point where they might eventually take on leadership roles. Then, colleges would receive more articulated resumes from applicants, while various school organizations would benefit stronger student participation, and students themselves would receive greater recognition for their efforts.

4. College Applications

Merely being accepted by a selective college or university requires much pre-planned effort, the time for which is literally unavailable to students already concerned with grades and other activities.

5. Grades

Struggling to achieve leadership roles and become the well-rounded applicants that colleges prefer, many

students find it difficult to maintain respectable grade-point-average during high school while remaining involved in activities that interest them. However, this tradeoff will not consider when colleges seek applicants with high grade-point-averages in their admissions pool. Lengthening the period of time spent in high school would allow more students to achieve both higher grades and fuller participation in extra-curricular activities on their lifeline. Rather than being forced to pick and choose among classes with the intention of evaluating their GPA, students could spread out their studies and make the most of every single year. With less pressure and more time, dedicated students could thus improve their grades. They would also enjoy studying these difficult subjects and interested in them. Some dropout students complete high school by means of an equivalency test and receive an alternative credential such as a General Educational Development certificate.

6. Individuality

Aside from the educational benefits, many students find that they shall be provided more time to accomplish a variety of goals. For example, while still at school, students will acquire a job so that the money earned might help them pay their way through college. With such a short preparation period before college, students can hardly be expected to accumulate an appropriate amount of funds for this purpose. In addition, many students who are interested in performing community service prior to attending college do not have enough time for this during the regular four-year high school program. Colleges will prefer students who have a rich background in community service and can show evidence of personal responsibility such as holding a job; however, under current circumstances, it remains difficult for students to unbalance their workload, and find the hours to put into these tasks.

7. Technical Skills

With classes, extracurricular activities, and part-time work, time is fleeting high school students. They have very little time to spend with their families, or for leisure activities. Even when some time will be available to recuperate, they often devote these extra hours to trying to accomplish more. Teens need to do what is necessary for their mental and physical health. They need to learn how to process effectively. Given the current pace many students' lives, this generation is likely to be full of workaholics, who do not understand or value the importance of family relationships.

8. Dropout Rates

Maintaining interest in their schoolwork is a trivial factor, if students are to succeed academically. Some high schools students drop out because they lose interest in school. Extending high school to five years in the way suggested here would reduce the number of students who drop out.

9. Conclusion

High schools lay the foundation for the rest of our life. Akin to money in the bank, the investment of an additional year when we're young can make an enormous difference. With the additional time that a five-year high school program would provide, motivated students would be able to become more involved in their studies, boost their grades, become gainfully employed, and engage in community service. Since colleges seek exactly these properties in successful applications for admission, high school students should therefore be provided with more time to take part in such endeavors. Students are clearly working hard to ensure their futures, and another year of high school would enable them to achieve this more effectively.

Document 2: DocL

Technology and the Lack of Leisure Time

1. Introduction

In a society where most households have clocks, phones, and televisions in almost every room, and daily schedules are demarcated by minutes instead of hours, many people suffer from stress and constantly whine that they don't have enough time to do everything they want to do. This complaint appears paradoxical when one considers the almost exponential development in technology. Computers are speedier and more powerful, with machines that do various tasks for us, while our means of transport takes us where we need to be much more quickly than in the past. Still, we have less time than ever to spend leisurely. This problem arises, not because technologies have failed to achieve the goal of improving efficiency, but rather because it has become a pursuit in itself, with our consumer society subjecting us to a basic ethical drive for "more".

2. Relaxing Pursuits

While technological advances have made many things better and faster, they have also created more activities. We now have televisions, computers, palm pilots, stereos, DVDs, play stations, and cell phones to occupy our time. Furthermore, all of these things are within reach of the average Canadians. Meanwhile, pursuits that consider as relaxing are becoming more expensive and less accessible. For example, for most people relaxing massages are affordable, and as cities grow larger, nature walks are becoming harder to found. It is sometimes easier for people to just sit down in front of the TV than to do something loose.

3. Diversity

Many Canadians complain that they do not have enough time. Most of us are caught up in a schedule of going to work each day, coming home late, then taking care of mandate details before finally falling into bed, only to get up early the next day to continue the routine. In most households, both parents work full-time, and are busy working throughout the week, which leaves them only the weekends to take care household duties such as cleaning and paying bills. As a result, many parents will feel ostracized by their children because the parents do not have enough time or energy to spend with their kids. Many people suffer from chronic stress because they do not take time out their busy lives simply to relax.

4. Effectiveness of Technology

Our computers are faster and more powerful, we have more machines that do various tasks for us, and our means of transport takes us to our destination much more quickly.

5. Intrusion of Media

What little free time people do have, they spend in front of the TV or on the computer, thus they are not relaxing, but are actively engaged in a cognitive process. Rather than relaxing, people experience tension or stress, because they have spent their entire time keeping up with the constant glimpse of images, storing information about characters, plots, themes, products and music, without realizing it. The intrusive aspect of media represents another reason why people do not have enough time. People fill their time with meditated technologies and caught up in their favorite shows or games rather than taking the time to perform other things that they have been "meaning to do". The use of email has been shown to eliminate a lot of visual and verbal cues we often use in communicating with one another.

6. Diversity

Canada was founded on the foundation of a conservative Protestant ethic which dictates that people shall work hard now so they may reap the rewards in the future. While this ethic is essential for productivity and pushes one to want more, it has become psychologically detrimental. People will push themselves to become more efficient so that they can accomplish more. However, there is the inclusion of more goals, thus, in effect, the work needed to achieve a goal never gets done. For example, people desire to buy more things to achieve higher social status for themselves. Hence, instead of saving for early retirement, they end up having to work longer to pay off their debts. In addition, people are so busy pushing themselves daily with the vague promise of retirement at the end, hoping to retire sooner in life that they will forget to stop and enjoy life as it is happening. In order to avoid working when they are older, and enjoy each day of their lives, people forgo daily enjoyments for the promise of future rewards.

7. Technology Addiction

"Leisure" should be a matter of personal choice, not filling time made available technologically advanced, efficient tools. The time saved by efficient tools will not necessarily transformed into leisure time. Ultimately, a person's motivation, in addition to his or her personality and lifestyle choices determine how the time saved is used. It is up to individuals to award the prize of leisure. They can choose to use it leisure time or for putting in more hours to make more money to support living expenses, or to buy the new car which they absolutely must have.

8. Social Impacts

Social interactions among people and the ways in which they communicate are affected by technology. Hence, there is no doubt that technological change prevents social change. Technologies leave little time for socialization and leisure.

9. Conclusion

While advances in technologies have made production more efficient, it has also burdened us with more things to manage and accomplish. As a result, people find themselves running around endlessly, sometimes forgetting what they're seeking. The point of any leisure time should be not to demand a greater affinity and reliance to technology, but a refusal to allow technologies run our lives. We must learn to stop occasionally, breathe, and enjoy life as it comes. We must learn to achieve a balance between looking ahead to tomorrow and learning to enjoy today.

C.3.3 Manipulations Planted in the Task Documents

This section contains the task documents with planted problems identified. The following keywords are used within the documents to indicate each of the problem category planted.

- *US*: Unclear Statements
- *FL*: Flawed Logic
- *TI*: Title Inconsistency
- *UA*: Unsupported Argument
- *GI*: Inconsistent Wording

- *AP*: Incorrect Use of Active-Passive Verbs
- *VT*: Incorrect Use of Verb Tense
- *MP*: Missing Prepositions
- *UC*: Use of Contractions
- *WC*: Inappropriate Word Choice

Document 1: DocH

Extending high school to five years

1. Introduction

"Senior itch" is an incessant irritation that high school students all crave to scratch. The cure? Graduation. As students progress through their high school years growing in wisdom and maturity, they are desperate for freedom. Yet what they desire most is not always what is best for them. Although most students don't[UC1] wish to admit it, extending the high school career to five years would make an important and beneficial affect[GI2] on their future. The current four years of high school education do not provide sufficient time for motivated students to accomplish all of their goals before college. Merely getting accepted by a selective college or university requires much pre-planned effort, and students who are preoccupied with their studies and other activities simply do not have enough time.

2. Interest

High schools are always adding new and interesting courses to the curricula which they offer. However, many college-bound students do not have sufficient time to explore these different classes. High school students thus are forced to choose classes that fit into their limited schedules. Occasionally classes that perceive[AI3] to be easier are chosen, merely to make sure that their transcripts stand out in the admissions pool. Fifteen years of age seems too late to start making choices based on "career"[FL4] , rather than choosing classes that they might find interesting and fun. An extra year of high school would provide students with opportunities to explore different subjects and take classes that they interest[AI5] . Consequently, students would enjoy their education and develop more stakes[US6] in school.

3. Individuality[TI7]

Colleges thoroughly consider the ways in which an applicant has spent his or her four years of high school. Students who take on leadership roles, display dictation to an organization[US8] , and show other evidence of being well-rounded appeal most to elite educational institutions. Although some high school students may desire to take on leadership positions[MP9] numerous extra-curricular organizations, they are often faced with school regulations that limit the number of offices they may hold. All too often, students will[VT10] find that they cannot participate in as many extra-curricular activities as they would like, due to the fact that they just do not have enough time. An extra year of high school would enable such students to become involved in more than one activity. Then, colleges would receive more articulated[WC11] resumes from applicants, while various school organizations would benefit stronger[MP12] student participation, and students themselves would receive greater recognition for their efforts.

4. College Applications

Merely being accepted by a selective college or university requires much pre-planned effort, the time for which is literally unavailable to students already concerned with grades and other activities.[UA13]

5. Grades

Struggling to achieve leadership roles and become the well-rounded applicants that colleges prefer, many

students find it difficult to maintain respectable grade-point-average during high school while remaining involved in activities that interest them. However, this tradeoff will not consider[AP14] when colleges seek applicants with high grade-point-averages in their admissions pool. Lengthening the period of time spent in high school would allow more students to achieve both higher grades and fuller participation in extra-curricular activities on their lifeline[US15] . Rather than being forced to pick and choose among classes with the intention of evaluating their GPA[US16] , students could spread out their studies and make the most of every single year. With less pressure and more time, dedicated students could thus improve their grades. They would also enjoy studying these difficult subjects and interested[AP17] in them. Some dropout students complete high school by means of an equivalency test and receive an alternative credential such as a General Educational Development certificate.[FL18]

-

6. Individuality[T119]

Aside from the educational benefits, many students find that they shall[VT20] be provided more time to accomplish a variety of goals. For example, while still at school, students will[VT21] acquire a job so that the money earned might help them pay their way through college. In addition, many students who are interested in performing community service prior to attending college do not have enough time for this during the regular four-year high school program. Colleges will[VT22] prefer students who have a rich background in community service and can show evidence of personal responsibility such as holding a job; however, under current circumstances, it remains difficult for students to unbalance their workload[US23] , and find the hours to put into these tasks.

-

7. Technical Skills[T124]

With classes, extracurricular activities, and part-time work, time is fleeting high[MP25] school students. Even when some time will[VT26] be available to recuperate, they often devote these extra hours to trying to accomplish more. Teens need to do what is necessary for their mental and physical health. They need to learn how to process effectively[US27] . Given the current pace many[MP28] students' lives, this generation is likely to be full of workaholics, who do not understand or value the importance of family relationships.

-

8. Dropout Rates

Maintaining interest in their schoolwork is a trivial factor[FL29] , if students are to succeed academically. Some high schools students [GI30] drop out because they lose interest in school. Extending high school to five years in the way suggested here would reduce the number of students who drop out. [UA31]

-

9. Conclusion

High schools lay the foundation for the rest of our life. Akin to money in the bank, the investment of an additional year when we're[UC32] young can make an enormous difference. With the additional time that a five-year high school program would provide, motivated students would be able to become more involved in their studies, boost their grades, become gainfully employed, and engage in community service. Since colleges seek exactly these properties[WC33] in successful applications for admission, high school students should therefore be provided with more time to take part in such endeavors.

Document 2: DocL

Technology and the Lack of Leisure Time

1. Introduction

In a society where most households have clocks, phones, and televisions in almost every room, and daily schedules are demarcated by minutes instead of hours, many people suffer from stress and constantly whine that they don't[UC1] have enough time to do everything they want to do. This complaint appears paradoxical when one considers the almost exponential development in technology. Computers are speedier and more powerful, with machines that do various tasks for us, while our means of transport takes us where we need to be much more quickly than in the past. Still, we have less time than ever to spend leisurely[GI2] . This problem arises, not because technologies have failed to achieve the goal of improving efficiency, but rather because it has become a pursuit in itself, with our consumer society subjecting us to a basic ethical drive for "more".

2. Relaxing Pursuits

While technological advances have made many things better and faster, they have also created more activities. We now have televisions, computers, palm pilots, stereos, DVDs, play stations, and cell phones to occupy our time. Furthermore, all of these things are within reach of the average Canadians. Meanwhile, pursuits that consider[AP3] as relaxing are becoming more expensive and less accessible. For example, for most people relaxing massages are affordable[FL4] , and as cities grow larger, nature walks are becoming harder to found[AP5] . It is sometimes easier for people to just sit down in front of the TV than to do something loose.[US6]

3. Diversity[TI7]

Many Canadians complain that they do not have enough time. Most of us are caught up in a schedule of going to work each day, coming home late, then taking care of mandate details before finally falling into bed,[US8] only to get up early the next day to continue the routine. In most households, both parents work full-time, and are busy working throughout the week, which leaves them only the weekends to take care household duties[MP9] such as cleaning and paying bills. As a result, many parents will[VT10] feel ostracized[WC11] by their children because the parents do not have enough time or energy to spend with their kids. Many people suffer from chronic stress because they do not take time out their busy lives[MP12] simply to relax.

4. Effectiveness of Technology

Our computers are faster and more powerful, we have more machines that do various tasks for us, and our means of transport takes us to our destination much more quickly.[UA13]

5. Intrusion of Media

What little free time people do have, they spend in front of the TV or on the computer, thus they are not relaxing, but are actively engaged in a cognitive process. Rather than relaxing[AP14] , people experience tension or stress, because they have spent their entire time keeping up with the constant glimpse of

images,[US15] storing information about characters, plots, themes, products and music, without realizing it. The intrusive aspect of media represents another reason why people do not have enough time. People fill their time with meditated technologies[US16] and caught[AP17] up in their favorite shows or games rather than taking the time to perform other things that they have been "meaning to do". The use of email has been shown to eliminate a lot of visual and verbal cues we often use in communicating with one another.[FL18]

-

6. Diversity[TI19]

Canada was founded on the foundation of a conservative Protestant ethic which dictates that people shall[VT20] work hard now so they may reap the rewards in the future. While this ethic is essential for productivity and pushes one to want more, it has become psychologically detrimental. People will[VT21] push themselves to become more efficient so that they can accomplish more. However, there is the inclusion of more goals, thus, in effect, the work needed to achieve a goal never gets done. For example, people desire to buy more things to achieve higher social status for themselves. Hence, instead of saving for early retirement, they end up having to work longer to pay off their debts. In addition, people are so busy pushing themselves daily with the vague promise of retirement at the end, hoping to retire sooner in life that they will[VT22] forget to stop and enjoy life as it is happening. In order to avoid working when they are older, and enjoy each day of their lives,[US23] people forgo daily enjoyments for the promise of future rewards.

-

7. Technology Addiction[TI24]

"Leisure" should be a matter of personal choice, not filling time made available [MP25] technologically advanced, efficient tools. The time saved by efficient tools will[VT26] not necessarily transformed into leisure time. Ultimately, a person's motivation, in addition to his or her personality and lifestyle choices determine how the time saved is used. It is up to individuals to award the prize of leisure.[US27] They can choose to use it leisure[MP28] time or for putting in more hours to make more money to support living expenses, or to buy the new car which they absolutely must have.

-

8. Social Impacts

Social interactions among people and the ways in which they communicate are affected by technology. Hence, there is no doubt that technological change prevents social change.[FL29] Technologies[GI30] leave little time for socialization and leisure. [FL31]

-

9. Conclusion

While advances in technologies have made production more efficient, it has also burdened us with more things to manage and accomplish. As a result, people find themselves running around endlessly, sometimes forgetting what they're[UC32] seeking. The point[WC33] of any leisure time should be not to demand a greater affinity and reliance to technology, but a refusal to allow technologies run our lives. We must learn to stop occasionally, breathe, and enjoy life as it comes. We must learn to achieve a balance between looking ahead to tomorrow and learning to enjoy today.

C.4 Tasks

C.4.1 Training Tasks

For Bundle System

Practice Tasks

1. Edit the first paragraph so that it will read as follows (the text needed to be inserted, deleted and replaced are indicated with bold font):
   *Some parenting experts urge **that** parents ~~**to**~~ tell their children the truth about their past ~~**experiment**~~**experience** with drugs and sex, even when they don't want their children to do what they did. Others think parents should censor what they tell their children. This paper will demonstrate that it is better for parents to censor what they tell their children until the children are old **and mature** enough to understand.*

2. Create a comment annotation with a tag, anchored to the phrase *"This paper"* (paragraph 1, last sentence) as follows:
   Comment text: ***Shall we say "We will" instead of "This paper"?***
   Tag: ***wordchoice***

3. Create a comment annotation as a follow-up comment on the *Insert* annotation *"and mature"* (paragraph 1, last sentence) as follows:
   Comment text:
   I want to include "mature" in addition to the word "old". I believe maturity is more important than the age to understand certain things in life.

4. Create a comment annotation, anchored to the phrase *"are developed"* (paragraph 2, second sentence) as follows:
   Comment text:
   Will the word "evolved" be more appropriate than "developed" here?

5. Edit the second last sentence in the second paragraph so that it will read as follows (replacements are indicated with bold font):
   *For instance, a woman who ~~**experienced**~~**experimented** with drugs as a teenager and ~~**regrets**~~**regretted** it will probably have developed the moral view that teenagers should not take drugs, based on (not despite) the fact that she did it herself.*

6. Create a new bundle named ***"Experiment vs. Experience"***, and add the annotations that replaced "experience" with "experiment" (i.e., annotations created in task 1 and 5). Add the following note to the bundle:
   Note:
   We are trying to say parents tried drugs as an experiment, aren't we?

7. Create a comment annotation with a tag, anchored to the phrase *"Children and young teens just don't get it"* (paragraph 2, last sentence) as follows:
   Comment text: ***Watch the tone here.***
   Tag: ***tone***

8. Create a comment annotation, anchored to the word *"hide"* (paragraph 2, last sentence) as follows:
   Comment text:
   This contradicts the sentence logically. I think you meant to say "reveal".

9. Create a new bundle named ***"Critical comments"***, and add the comment annotation that you created in the previous task (*"This contradicts the sentence logically..."*).

10. Switch to "All Annotations" tab. Select the annotation that you created in the task 3: (*"I want to include mature in addition to the word "old"..."*). Add the selected annotation to the bundle you created in the previous task *"Critical comments"* by clicking "Add to Bundles" button.

11. Edit the comment you created in the task 4 (*"Will the word "evolved" be more appropriate..."*) by adding a tag named ***"wordchoice"***.

12. From *All Annotations,* filter the *comment* annotations that have the tags *"wordchoice" or "tone"*. Save the filter results in a bundle named ***"Need rewrite"***.

13. Remove the comment annotation: *"Shall we say 'We will' instead of 'This paper'* from the bundle named *"Need rewrite"*. Then save the bundle to reflect the changes, and close the bundle tab.

14. To which bundles (if any) are associated with the annotation *"Replace: experiment with experience"* from the first paragraph. (Hint: you can use "View Associated Bundles" from the right click menu to check)

15. Delete the insert annotation *"Inserted: and mature"* from the document. (Hint: you might notice that you will get an error saying you can't delete nested annotations. Try to delete the follow-up comment of *"I want to include 'mature' in addition..."* before deleting the insert annotation.)

16. Create a general comment as follows:
    Comment text:
    This paper seems on track. But you might want to add more supporting evidence for your argument.

For Simple System

Practice Tasks

1. Edit the first paragraph so that it will read as follows (the text needed to be inserted, deleted and replaced are indicated with bold font):

   Some parenting experts urge ***that*** *parents* ~~***to***~~ *tell their children the truth about their past* ~~***experiment***~~***experience*** *with drugs and sex, even when they don't want their children to do what they did. Others think parents should censor what they tell their children. This paper will demonstrate that it is better for parents to censor what they tell their children until the children are old* ***and mature*** *enough to understand.*

2. Create a comment annotation, anchored to the phrase "*This paper*" (paragraph 1, last sentence) as follows:

   Comment text: ***Shall we say "We will" instead of "This paper"?***

3. Create a comment annotation, anchored to the inserted words "*and mature*" (paragraph 1, last sentence) as follows:

   Comment text:

   I want to include "mature" in addition to the word "old". I believe maturity is more important than the age to understand certain things in life.

4. Create a comment annotation, anchored to the phrase "*are developed*" (paragraph 2, second sentence) as follows:

   Comment text:

   Will the word "evolved" be more appropriate than "developed" here?

5. Edit the second last sentence in the second paragraph so that it will read as follows (replacements are indicated with bold font):

   For instance, a woman who ~~***experienced***~~***experimented*** *with drugs as a teenager and* ~~***regrets***~~***regretted*** *it will probably have developed the moral view that teenagers should not take drugs, based on (not despite) the fact that she did it herself.*

6. Create a comment annotation, anchored to the phrase "*Children and young teens just don't get it*" (paragraph 2, last sentence) as follows:

   Comment text: ***Watch the tone here.***

7. Create a comment annotation, anchored to the word "*hide*" (paragraph 2, last sentence) as follows:

   Comment text:

   This contradicts the sentence logically. I think you meant to say "reveal".

8. From *All Annotations,* filter the *comment* annotations that have been created.

9. Edit the comment you created in the task 4 (*"Will the word "evolved" be more appropriate..."*) by adding a phrase saying ***"since we are arguing that parents' views and values are changing over time?"***

10. Delete the comment annotation you created in task 3 *"I want to include "mature" in addition..."* from the document. Then delete the insert annotation *"and mature"* from the document.

C.4.2 Experiment Tasks

For DocH

Scenario:

You have been recruited to participate in this study because you have expertise in reviewing documents and helping students improve their documents' quality and their writing skills. In this study, you will be asked to take a similar role, whose task involves reviewing documents.

You will be playing the role of a *co-author* who is involved in a collaborative writing project that will result in a published paper (i.e., your name will be included as one of the co-authors for the publication). Your responsibilities in this project include reviewing the documents to improve the quality, and assisting your co-authors with the writing process.

Imagine you are Sam, and you are collaborating with two other co-authors: Mary and John on an argumentative article, titled "Extending High School to Five Years". Mary and John are the primary authors of the given document and have written a complete draft of the document. They have divided the document into sections and each wrote different sections as followings (note that the paragraphs are numbered in the document as well):

- **1st and 9th paragraphs are co-written by both Mary and John**
- **3rd, 6th, 7th paragraphs are written by Mary**
- **2nd, 4th, 5th and 8th paragraphs are written by John**

You are about to leave town for summer vacation tomorrow. They have requested that you review the document before you leave. Upon receiving your review, they will read your annotations and incorporate them into the document.

Your task is to review the document; you can make **direct edit changes** to the document text or **provide comments**. Since you are not a primary writer in this project, you do not need to rewrite the document, but rather you need to point out the problems and/or make suggestions in the document.

Along with your review, your co-authors requested that you provide the following specific feedback, but not limited to:

- your general impression of the writing
- brief summary of your review that provides an overall status of the document which will help your co-authors skim the document quickly and prioritize the remaining work.

Please note that you can communicate your co-authors *only* via the tool(s) provided. The deadline for submitting the document is during your vacation period. They will not be able to access you before the submission once you leave tomorrow. Hence, please make sure that you complete the task (i.e., you need to **finish reviewing the *whole* document** and provide **clear annotations** along with **responses to their requests.**)

You have **an hour** to complete the task. Please try to think out loud or narrate your thoughts while performing the task.

For DocL

Scenario:

You have been recruited to participate in this study because you have expertise in reviewing documents and helping students improve their documents' quality and their writing skills. In this study, you will be asked to take a similar role, whose task involves reviewing documents.

You will be playing the role of a *co-author* who is involved in a collaborative writing project that will result in a published paper (i.e., your name will be included as one of the co-authors for the publication). Your responsibilities in this project include reviewing the documents to improve the quality, and assisting your co-authors with the writing process.

Imagine you are Sam, and you are collaborating with two other co-authors: Mary and John on an argumentative article, titled "Technology and the Lack of Leisure Time". Mary and John are the primary authors of the given document and have written a complete draft of the document. They have divided the document into sections and each wrote different sections as followings (note that the paragraphs are numbered in the document as well):

- **1st and 9th paragraphs are co-written by both Mary and John**
- **3rd, 6th, 7th paragraphs are written by Mary**
- **2nd, 4th, 5th and 8th paragraphs are written by John**

You are about to leave town for summer vacation tomorrow. They have requested that you review the document before you leave. Upon receiving your review, they will read your annotations and incorporate them into the document.

Your task is to review the document; you can make **direct edit changes** to the document text or **provide comments**. Since you are not a primary writer in this project, you do not need to rewrite the document, but rather you need to point out the problems and/or make suggestions in the document.

Along with your review, your co-authors requested that you provide the following specific feedback, but not limited to:

- your general impression of the writing
- brief summary of your review that provides an overall status of the document which will help your co-authors skim the document quickly and prioritize the remaining work.

Please note that you can communicate your co-authors *only* via the tool(s) provided. The deadline for submitting the document is during your vacation period. They will not be able to access you before the submission once you leave tomorrow. Hence, please make sure that you complete the task (i.e., you need to **finish reviewing the *whole* document** and provide **clear annotations** along with **responses to their requests.**)

You have **an hour** to complete the task. Please try to think out loud or narrate your thoughts while performing the task.

Appendix D

UBC Research Ethics Board Certificate

The University of British Columbia
Office of Research Services
Behavioural Research Ethics Board
Suite 102, 6190 Agronomy Road, Vancouver, B.C. V6T 1Z3

CERTIFICATE OF APPROVAL- MINIMAL RISK RENEWAL

PRINCIPAL INVESTIGATOR:	DEPARTMENT:	UBC BREB NUMBER:
Kellogg S. Booth		H05-80494

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT:

Institution	Site
UBC	Point Grey Site

Other locations where the research will be conducted:
N/A

CO-INVESTIGATOR(S):

Joanna McGrenere
Yamin Htun

SPONSORING AGENCIES:

Natural Sciences and Engineering Research Council of Canada (NSERC) - "Network for Effective Collaboration Technology through Advanced Research"

PROJECT TITLE:
B05-0494 - Network for Effective Collaboration Technology through Advanced Research

EXPIRY DATE OF THIS APPROVAL: June 5, 2008

APPROVAL DATE: June 5, 2007

The Annual Renewal for Study have been reviewed and the procedures were found to be acceptable on ethical grounds for research involving human subjects.

Approval is issued on behalf of the Behavioural Research Ethics Board