

3
2008



This is to certify that the
dissertation entitled


Hypermedia and learning:
Contrasting interfaces to
hypermedia systems

presented by

Amy Tracy Wells

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Counseling Educational
Psychology, and Special
Education


Major Professor's Signature

5/7/08
Date

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

**HYPERMEDIA AND LEARNING: CONTRASTING INTERFACES TO
HYPERMEDIA SYSTEMS**

By

Amy Tracy Wells

A DISSERTATION

**Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of**

DOCTOR OF PHILOSOPHY

Department of Counseling Educational Psychology, and Special Education

2008

ABSTRACT

HYPERMEDIA AND LEARNING: CONTRASTING INTERFACES TO HYPERMEDIA SYSTEMS

By

Amy Tracy Wells

This study explores selected theoretical and design issues associated with the use of hypermedia learning environments to promote the recall, synthesis, integration and retention of information. Specifically, the study contrasts two different hypermedia systems that contain resources on the Flint Sit-Down Strike, a complex historical domain. The experimental condition incorporates design features related to complexity, context-dependency and interconnectedness in order to highlight different aspects of its instructional content. The experimental condition was hypothesized to foster greater achievement on tests for synthesis, integration and retention of knowledge and to be more favorably rated by users. The control condition incorporates simple linear design features interface including several features that are antithetical to those of the experimental condition. The control condition presents the same instructional content in a more rigid and decontextualized manner and was hypothesized to foster greater mastery of factual recall but less synthesis, integration and retention of knowledge. Results however demonstrated that participants in the control condition were able to recall more facts, make more connections between themes in the test for synthesis and retain more facts in the test for retention than participants in the experimental condition. However, differences in *overall*

performance for both hypermedia systems were not statistically significant as there was no difference in the number of facts cited in the test for factual integration or synthesis. Lastly, there was no significant difference in overall performance between the two conditions on the integration test.

This study's major contributions include (1) a methodology for comparing and testing interfaces, (2) the finding of no difference in overall performance between the two interfaces and (3) a review of the literature, and a series of alternative conclusions and recommendations for future research.

Copyright by

AMY TRACY WELLS

2008

Dale, this is not for you, it is because of you.

In loving memory of Clare D. and Albert A. Belman, and Violet Wells for always,
always encouraging me

In loving memory of Gerald R. Wells

Thank you to my children, Livia and Aaron, for sharing me.

ACKNOWLEDGEMENTS

I am grateful for the intellectual and moral support shown to me by many in the Program but most especially my dissertation committee: Drs. Patrick Dickson, Raven McCrory, Dean Rehberger, Rand Spiro and Victor Rosenberg. I would like to especially thank Rand Spiro for his guidance throughout my graduate school years and to Raven McCrory for her assistance in turning this research into a dissertation.

I am also indebted to Dr. Mark Kornbluh and MATRIX staff for providing the expertise and computing resources necessary to the development and testing of the respective systems, which was funded by a grant from the Digital Libraries Initiative II: Digital Libraries in the Classroom Program—Joint Information Systems Committee and National Science Foundation, award no. IIS-0229808.

TABLE OF CONTENTS

LIST OF TABLES.....	x
LIST OF FIGURES	xii
CHAPTER 1. INTRODUCTION	1
Online Library Catalogs	2
Digital Libraries	5
Conceptual Framework	9
Overview of Study and Experimental Hypotheses	12
Limitations.....	20
CHAPTER 2. LITERATURE REVIEW	21
Relevant Empirical Literature.....	22
Individual Differences	25
Structural Differences	27
Efficiencies	29
Satisfaction, Attitude and Motivation	30
Summary	31
Cognitive Flexibility Theory	32
Why is this research important?	36
CHAPTER 3. RESEARCH METHODS	39
Testing and Learning Stages	39
Background and Materials	42
Sample	45
Heuristic Testing	48
Conditions	55
Pilot Testing	67
Lab and Equipment.....	70
Testing Procedures.....	70
Procedural Overview	70

Content Knowledge	72
Evaluation Measures	73
Recall, Synthesis and Integration.....	74
Attitude survey	76
Vocabulary Testing	77
Personal Data	78
Retention Measure.....	78
Scoring Procedures	80
Statistical Tests	87
 CHAPTER 4. STATISTICAL ANALYSIS	 88
Survey I	89
Background Performance Measures.....	90
Content Knowledge	90
Vocabulary	93
Personal information	93
Attitude	94
Recall.....	106
Descriptions of Essays	109
Integration	111
Synthesis	119
Survey II: Post test	121
Content Knowledge	122
Retention	125
Implications of Sample Size for these Findings.....	127
Did Learning Occur?.....	128
Did the Systems have Distinct Effects on Learning?	130
Summary	134
 CHAPTER 5. ANALYSIS.....	 138
How does this research relate to the hypermedia literature?	138
What does this research say about Cognitive Flexibility Theory?	138

How does this research change our understanding of hypermedia
systems? 139

Alternative conclusions..... 143

Recommendations for further research 146

APPENDICES 149

REFERENCES: 221

LIST OF TABLES

Table 1-1. Summary of the Major Features of the Control and Experimental Groups.....	14
Table 2-1. Cognitive Flexibility Principles.....	34
Table 3-1. Summary of Variables.....	44
Table 3-2. Summary of Participants	47
Table 3-3. Phases in the Heuristic Evaluation.....	49
Table 3-4. Summary of the Heuristic Evaluation – components and procedure	51
Table 3-5. A breakdown of the audio files, images and texts in both conditions by theme area.....	62
Table 3-6. A numeric breakdown of total and unique resources.....	64
Table 3-7. A numeric breakdown of unique and shared resources by theme area.....	64
Table 3-8. Number of words in Text Files	66
Table 3-9. Word count for introductory paragraphs	66
Table 3-10. Time Required	72
Table 3-11. Synthesis Measure Rubric	84
Table 3-12. Retention Measure Rubric	86
Table 4-1. Regression on Attitudes	100
Table 4-2. Rank Ordering of Responses to “The best part of the (Control) system was...”	102
Table 4-3. Rank Ordering of Responses to “The best part of the (Experimental) system was...”	103
Table 4-4. Rank Ordering of Responses to “The most difficult part of the (control) system was...”	104
Table 4-5. Rank Ordering of Responses to “The most difficult part of the (experiment) system was...”	105
Table 4-6. Regression on Recall.....	109
Table 4-7. Overall means.....	112
Table 4-8. Integration Rubric	113
Table 4-9. Integration Essay	114
Table 4-10. Regression on Total Integration Score	116
Table 4-11. Regression on Number of Themes	117
Table 4-12. Regression on Number of Facts	118
Table 4-13. Regression on Connections	119
Table 4-14. Synthesis scores by condition	120
Table 4-15. Regression on Total Synthesis	121
Table 4-16. Content knowledge	122
Table 4-17. True/False Correct	122
Table 4-18. Labor Leaders Correct.....	123

Table 4-19. Regression of T/F Post Test	124
Table 4-20. Regression of Labor Leaders Post test.....	125
Table 4-21. Retention Scores	126
Table 4-22. Regression of Total Retention.....	127
Table 4-23. Descriptive Statistics Pre-test to Post-test	129
Table 4-24. T/F Experimental and Control Condition	129
Table 4-25. Labor Leaders Experiment and Control Condition	130
Table 4-26. Differences in T/F and Labor Leader responses.....	131
Table 4-27. Differences in T/F responses*	131
Table 4-28. Differences in Labor Leader responses*	132
Table 4-29. Regression of T/F Questions	133
Table 4-30. Regression of Labor Leaders Questions.....	134
Table 4-31. Summary of Difference by Condition	136

LIST OF FIGURES

Images in this dissertation are presented in color.

Figure 1-1. Online Library Catalog from the Library of Congress	3
Figure 1-.2. Digital Library from the Library of Congress.....	6
Figure 1-3. Control Condition Site Map	17
Figure 1-4. Experimental Condition Site Map.....	18
Figure 3-1. Linear/ control condition	57
Figure 3-2. Linear/ control bibliographic record	58
Figure 3-3. Context-rich/ experimental condition	60
Figure 3-4. Complex rich/ experimental resource.....	61
Figure 3-5. Distribution of the length of sound files	65
Figure 3-6. Pilot Test Testing Flowchart.....	68
Figure 3-7. Integration Measure Rubric	82
Figure 4-1. Attitude toward the System (1)	97
Figure 4-2. Attitude toward the System (2)	98

CHAPTER 1. INTRODUCTION

The use of hypermedia¹ systems and resources for learning has opened a series of questions about which hypermedia designs most effectively support learning. Do systems that allow users to browse contextualized materials and create implicit connections among resources promote learning more effectively than more linear systems that compel users to search and retrieve discrete resources? Do these different systems promote different types of learning such as recall or understanding? Although hypermedia and its effect on learning has been researched along several dimensions including how different designs affect learning; to-date no empirical research has examined two dominant variants of information systems, the online library catalog and digital libraries. This research examines how different interfaces that simulate an online library catalog and digital library² affect learning, specifically the factual recall, synthesis, integration and retention of information.

The research uses an experimental design to study the effects of contrasting displays of hypermedia information on learning. The first condition, the control, uses a linear hypermedia design that replicates the functions and capabilities³ of the Library of Congress' Online Catalog and Michigan State University's MAGIC online catalog. As such, the control is a realistic and common technology though the content in this case links to full text, audio and image files

¹ Hypermedia refers to systems composed of different forms of media such as audio, images and text that are linked together.

² For the sake of simplicity this dissertation will use the term "library catalog" to refer to the linear interface and "alternative digital library interface" (ADLI) to refer to the non-linear and more complex, context-rich hypermedia interface.

³ Current as of March 2006.

and contains a limited number of resources. The second condition, the experimental, uses a more complex, context-rich hypermedia design and replicates the general capabilities of some alternative digital library interfaces (ADLIs), though, as with the control, the content itself is limited. The general functionality that has been replicated in the experimental condition includes an ability to browse resources, the presence of textual, graphical and auditory cues and supplemental contextual information. Online library catalogs and digital libraries differ in many ways including the presentation and form of their content. (These differences are explained more fully in the sections that follow.) The intent of this research is to study how these contrasting interfaces influence the cognitive processes of the people who use these systems.

Online Library Catalogs

Online library catalogs⁴ (Figure 1-1. Online Library Catalog from the Library of Congress), which have emerged in the past twenty-five years, have enabled patrons to search the physical contents of academic, public and special libraries with unprecedented speed. Online library catalogs provide an interface for searching the contents of a library's collection. The collections typically include text in print and electronic forms, and resources in other formats such as maps, video, software, etc. Individuals type in a text string (e.g., "Shakespeare's sonnets") that is matched against indexes for the collection and a resulting display of items is returned. These "matches" are displayed in a list format.

⁴ This definition addresses characteristics found in library catalogs and is current as of January 2007. See also, for example, Thomas, 2000, for additional specifics.

Individuals can then click on any given item to learn more about an item including its availability. The majority of any library's collection is typically in physical form which then requires that each item be located and accessed individually and after some period of time has elapsed. A much smaller percentage of most libraries' collections is electronic, and access to these items may be direct depending on a series of conditions involving rights management.

The Library of Congress

LIBRARY OF CONGRESS ONLINE CATALOG

DATABASE: Library of Congress Online Catalog

Basic Search Guided Search

Search Text:

Search Type:

Author/Creator Browse

Subject Browse

Keyword

Command Keyword

Call Number Browse (LC Class)

Call Number Browse (Other)

LCCN-BN-HSSN

Search/Online Table Browse

Author/Creator Sorted by Title

25 records per page

Begin Search Clear Search Set Search Limits

Basic Search Tips

Search Type	Brief Help: Select a "search type" for detailed Help	Help on Limits
Title	<ul style="list-style-type: none"> Enter all or initial part of title, starting with the first word: king and i Drop initial articles (a, an, the) and punctuation (., &lt; >, !, ?) in any language. Truncation is automatic. 	Search Limits Available!
Author/Creator Browse	<ul style="list-style-type: none"> For personal names, enter surname first: shakespeare william For corporate bodies, enter in direct order: new york college Truncation is automatic. 	Search Limits Not Available!
Subject Browse	<ul style="list-style-type: none"> Enter standard subject heading, name, or genre term: social energy Start with the set-most word: vietnamese conflict 1963-1975 Clear most punctuation: united states army library (especially double hyphens in subject) 	Search Limits Not Available!

Figure 1-1. Online Library Catalog from the Library of Congress

Despite the utility of online catalogs, the catalogs themselves and their holdings remain difficult for many people to understand. Mimicking the traditional card catalogs, online library catalogs are typically designed so individuals can

perform searches for materials via title, author, keyword or subject heading and return syntactically or semantically related items. Individuals then examine these returned items and search again and/or retrieve the physical items, or access them electronically. While there is no one weak link in this process, much of the research focuses on the difficulties people have in querying systems using rigid, controlled vocabularies in the form of subject headings (Bates, 2003; Sandberg-Fox, 2001; Borgman, 1996; Drabenstott, 1991). A controlled vocabulary is a predefined list of terms, which, in this case, is used to describe resources. Its purpose is to group common resources. A simple example relates to the terms "attorney" and "lawyer." Instead of using one term or the other or both, a controlled vocabulary might require that only the term "attorney" be used to describe resources that might otherwise also be described using the term "lawyer". Ideally, a controlled vocabulary enables an individual searching the catalog to locate all conceptually related items. The process of locating information then becomes one of trying to identify an information need, which an individual may or may not be able to conceptualize, and articulating the need in terms defined in a controlled vocabulary (Borgman, 1996).

Research also indicates that there is a significant divide between those who understand subject headings (i.e., the controlled vocabulary that is actually used) in a library catalog function and those who do not (Drabenstott, 1991). So while any given item in a library's collection may be assigned and subsequently located under multiple subject headings, in practice the item may remain difficult to locate. Interestingly, many library patrons would prefer to browse the physical

collection pertinent to their area of interest rather than use subject headings or use the catalog at all (Bates, 2003). Following an extensive review of the information seeking literature, Bates (2003) concludes that "...browsing may in fact be the dominant and most natural form of searching, and that systems that make information discovery feel like browsing, whatever their actual structure, will attract more users and help those users to be more effective information seekers" (p. 14).

Digital Libraries

Digital libraries⁵ (Figure 1-.2. Digital Library from the Library of Congress), which have emerged in the past fifteen years and have become more and more prevalent, differ from online library catalogs in three significant ways. Perhaps most significantly, in a digital library the resources can be browsed⁶ via a computer. In addition, textual, graphical and auditory cues may also be present to guide information seeking. Digital libraries, which function both as a catalog to their resources and as an information space, vary tremendously in their design, underlying metadata structures and support. In further contrast to online library catalogs, digital libraries at present do not seek to cover all intellectual domains.⁷ Rather, they typically have a specific focus, such as American History. Digital

⁵ This definition addresses characteristics found in digital libraries and is current as of January 2007. See also Borgman, 2004, 2000 and 1999 for additional specifics.

⁶ The term "browsing" has many definitions within the field of Information Science (Rice, McCreddie & Chang, 2001). A common element of these definitions is the idea that browsing includes undirected or semi-directed searching or scanning (Bates, 2002; Bates, 1989; Ellis, 1989).

⁷ Though this dissertation focuses on free and freely accessible digital libraries, the preceding comments also apply to commercial digital libraries and archives such as the Association for Computing Machinery's (ACM) Portal, Corbis, *New York Times*, etc.

libraries do however provide individuals with a variety of media, including short essays, graphics, timelines, etc., which help to contextualize subjects, people and/or events.

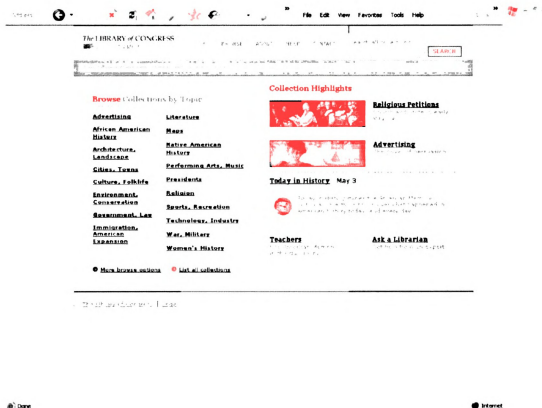


Figure 1-2. Digital Library from the Library of Congress

One of the best known and largest of the digital libraries, "American Memory," (Library of Congress) is focused on the experience of people living in America. Though the collection has continued to grow, it has been developed subject-area by subject-area (e.g., African-American History, Women's History, etc.). Its design has been heavily influenced by online library catalogs⁸. However, instead of requiring the individual to type out a subject heading such as "Suffragists--United States--1910-1920," topics and resources (such as the following) that answer or guide individuals' information seeking are displayed in a browsable, hierarchical format, which allows the individual to simply click on links of interest:

→ Women's History

→ Woman Suffrage ~ Photographs ~ 1875-1938

→ Women of Protest: Photographs from the Records of
the National Woman's Party

→ Brief Timeline of the National Woman's Party
1912-1997

This permits the individual to "drill down" to the resources much more easily. In addition, context about the resources is presented and the resource is accessible directly on the screen (or from a speaker) so the user does not have to retrieve the physical item to determine its usefulness for his or her purpose.

⁸ American Memory, which is comprised of different collections, developed at different times, contains collections that have features more associated with traditional catalogs while it also has collections that have features more associated with digital libraries.

Other examples of browsing designs include the National Science Digital Library (NSDL) (2007) and the Alexandria Digital Library (2007). The NSDL uses visualization software to group and relate concepts, but also offers a text-based interface to enable browsing. For example, individuals can search the NSDL using a specific term such as “communications technology” and then click on different subcategories of resources on communications technology (e.g., devices and media, coding and decoding, and quality) in order to retrieve links to the resources. The Alexandria Digital Library permits geospatial browsing as well as longitudinal/ latitudinal and temporal periods of organization. For example, individuals can click on “California,” then “San Diego” and choose different maps, air photos and satellite images of the area from 1950 to the present.

These three examples have several things in common. Each digital library provides nonlinear access to the resources, contains contextual information in the form of textual and visual cues that guide information seeking, demonstrates the relationships amongst resources and functions as self-contained information space by providing immediate access to its materials.

There are cognitive consequences to the general design of digital libraries. Though the formats vary (e.g., mp3, MS Word, Adobe Acrobat), the resources themselves, including image, audio, and motion picture, can be accessed online. As a result, the patron has almost immediate access to the resources, allowing for an uninterrupted thought process. Traditionally, information seeking, for even the most proficient patron, has required a time consuming physical component that disrupts the patron’s cognitive process (i.e., retrieving materials from the physical

library). It may be that the comparative simultaneity of seeking and retrieving resources offered by digital libraries results in the patron developing new cognitive connections between the question or information need and possible resources.

This research examined how these two different interfaces, a simulated online library catalog and a simulated digital library both of which contained electronic resources, affect cognitive processing. Participants were exposed to information in one of two interfaces and their factual recall, synthesis, integration and retention of information was assessed. The goal of this research was to determine the effects of system design on a range of processes and determine which, in the context of this study, better supported learning.

Conceptual Framework

A significant problem in the design of hypermedia systems and, more broadly, in information science and educational technology, is the lack of a conceptual framework to guide design and testing (Pettigrew, Fidel & Bruce, 2001; Mishra, Spiro & Feltovich, 1996; Spiro, Feltovich, Jacobson, & Coulson, 1992; Spiro & Jehng, 1990). This study seeks to understand how a hypermedia learning environment that employs two principles from Cognitive Flexibility Theory (CFT) affects cognitive processes in comparison to a simulated online library catalog. Specifically, context-dependency and interconnectedness, which acknowledge and manage complexity, have been used in the design of a simulated digital library but not in the design of the simulated online library catalog. In the next section, it is argued that although these principles appear to varying degrees in

different digital libraries, neither CFT nor its principles have been discussed in the context of digital library design.

CFT is a prominent theory of learning which bridges cognitive constructivist approaches and collaborative, sociocultural, and situated views to promote advanced knowledge acquisition (Spiro, Collins, Thota & Feltovich, 2004). CFT's focus is not on the content of learning environments, "...but rather how their *form* influences the cognitive structures and processes of those who use them" (Mishra et al., 1996, p. 17). Underlying the theory are principles that have been designed to promote non-insular understandings of complex information for their application in real-world contexts (Spiro et al., 2004; Mishra et al., 1996; Spiro, Feltovich, & Coulson, 1996; Jacobson & Spiro, 1995; Spiro & Jehng, 1990; Spiro, Vispoel, Schmitz, Samarapungavan & Boerger, 1987). CFT does not reject objective reality or deny an ability to capture reality but notes that knowledge that is to be used in multiple contexts must be presented along different conceptual dimensions in order to capture complexity. If not, representations are inadequate and lead to reductive understandings ultimately resulting in the learners' inability to apply knowledge successfully (Spiro et al., 1992; Spiro et al., 1987).

The CFT principles of context-dependency and interconnectedness have been used in the design of the experimental condition. The first of these two principles, context-dependency, refers to the idea that "facts do not remain self-evident, isolated bits of information but, rather, are 'constructed' by their perceived relationship to other facts and by their usefulness in understanding cases" (Mishra et al., 1996, p.8). The second of these two principles, interconnectedness, refers

to the idea that “conceptual and case knowledge cannot be ‘boxed’ into separate *mental* compartments...[but rather requires juxtaposition in order to support] the *goal* of widely applicable or transferable knowledge” (Spiro, Collins, Thota & Feltovich, 2004).

Most notable in the context of the present research is a study by Jacobson & Spiro (1995) that reports a significant effect in their experimental research for a *hypertext* system that used several principles from CFT including:

- using multiple conceptual representations of knowledge,
- linking and tailoring abstract concepts to different case examples,
- introducing domain complexity early,
- stressing the interrelated and web-like nature of knowledge, and
- encouraging knowledge assembly.

The control group(s) used a traditional drill-based design that focused on *presenting* a theme or themes with specific mini-cases as opposed to *demonstrating* the complexity of the themes and how they overlapped with various *mini*-cases. The first control group (I) was exposed to a series of mini-cases and a *singular theme*. A second control group (II) was exposed to the same mini-cases *but* with *multiple themes* highlighted, while the third, experimental group, was *exposed* to mini-cases, their associated themes with commentaries, and *thematic criss-crossing*. (Control groups I and II were later collapsed into a single control group as they were found to be statistically equivalent.) The control group *completed* their study stage sessions in a significantly shorter period of time and *were* more effective than the experimental group at acquiring factual knowledge.

However, the experimental group "...was found to have significantly higher adjusted mean problem-solving essay scores..." (p. 321) than the control group.

Overview of Study and Experimental Hypotheses

The experimental and control condition make use of electronic resources about the Flint Sit-Down Strike of 1936-1937, which was a strike of national importance. However, two different approaches to structuring the hypermedia resources (i.e., audio, short and long texts, and images) that cover "working conditions," "strike methodology," "conditions during the strike" and "community response" are used. Specifically, contextual information and interconnectedness, principles of CFT, which are often features of digital libraries and not of online library catalogs, are incorporated in the experimental condition.

The resources and resource descriptions (type of resource, title, author, description, date and subject headings); number of resources; and online accessibility of resources are the same in both the experimental and the control conditions. However, as Table 1-1. Summary of the Major Features of the Control and Experimental Groups indicates, the experimental condition incorporates design features related to complexity, context-dependency and interconnectedness in order to highlight different aspects of its instructional content. Specifically, the experimental condition contains brief contextual information that is designed to provide a minimal introduction to the hypermedia learning environment and its content. The map of Flint within Michigan and within the United States along with images of the people involved in the Flint Sit-Down Strike provide additional context. This contextual information, a principle of CFT, is

not present in the control condition. The contextual information includes one general introduction (49 words) to the collection, four specific introductions (69 to 93 words) to each of the themed-areas (i.e., “working conditions,” “strike methodology,” “conditions during the strike” and “community response”), and one graphic map that situates Flint in Michigan, and Michigan in the United States. Interconnectedness, which is highly related to context-dependency, is conveyed through the grouping of resources into themes and the subsequent inclusion of resources in multiple themes in order to demonstrate the connections between people and events. Key to demonstrating the interconnectedness of the people and events involved in the Flint Sit-Down Strike is the fact that resources themselves can be accessed quickly thereby helping individuals to form mental connections.

Table 1-1. Summary of the Major Features of the Control and Experimental Groups	
Linear hypermedia design/Control condition	
<p>Individuals must search for the information needed, locate a relevant bibliographic record and then click within the object to access the actual information. Each bibliographic record provides the type of information (e.g., text, audio and images), title, author, description, date and subject as well as a direct link to the information itself. The control condition is intended to present the same resources as the experimental condition, but in a rigid and decontextualized manner.</p>	
CFT element: Context-dependency	CFT element: Interconnectedness
No additional context (e.g., facts or circumstances) or media is presented.	Information (i.e., text, audio and images) is presented discretely with no interconnected (e.g., coordinated) content.
Context-rich hypermedia design/Experimental condition	
<p>Several presentations of the same information in multiple contexts were provided to highlight different facets of the information. In addition, the experimental condition displays an interwoven series of images, text and audio to form a context-rich representation. Specifically, individuals browse thematic areas (though they can also search as well), locate a related informational object, and automatically see or hear the information while discovering the type of information (e.g., text, audio, and images), title, author, description, date and “themed”</p>	

<p>Table 1.1 (cont'd).</p> <p>information. This condition is intended to present the same resources as in the control condition but in a cognitively flexible format that will enhance the ability of learners to acquire new knowledge and to transfer prior knowledge to new situations.</p>	
CFT element: Context-dependency	CFT element: Interconnectedness
<p>Contextual information including one general introduction (49 words) to the collection, four specific introductions (40 to 54 words) to each of the themed-areas (i.e., “working conditions,” “strike methodology,” “conditions during the strike” and “community response”) is presented.</p> <p>A graphical map of Flint within Michigan and within the United States along with images of the people involved in the Flint Sit-Down Strike provides additional context.</p>	<p>Interconnectedness is conveyed through the grouping of resources into themes and the subsequent inclusion of resources in multiple themes in order to demonstrate the connections between people and events.</p> <p>Key to demonstrating the interconnectedness of the people and events involved in the Flint Sit-Down Strike is the fact that resources themselves can be accessed quickly thereby helping individuals to form mental connections.</p>

Though online library catalogs do not typically contain access to the full content of all of their resources, in contrast to digital libraries, the full content is

available in the control condition. However, access to the resources in the control condition requires four clicks while access in the experimental condition requires three clicks. This replicates the functionality of online library catalogs, which typically require additional clicks to reach the resource itself in contrast to digital libraries.

In summary, the majority of the content and media are the same in each of the two systems. However, the information is organized and accessed differently. In simplest terms, individuals in the control condition had to search for information (Figure 1-3. Control Condition Site Map) while individuals in the experimental condition had to select information and/or resources while browsing (Figure 1-4. Experimental Condition Site Map). That is, in the control condition individuals had to search for the information needed, locate a relevant bibliographic record and then click within the object to access the actual information. Each bibliographic record provided the type of information (e.g., text, audio and images), title, author, description, date and subject as well as a direct link to the information itself. The control condition was intended to present the *same resources* as the experimental condition, but in a rigid and decontextualized manner. In contrast, the experimental condition displayed an interwoven series of images, text and audio to form a context-rich representation. Specifically, individuals browsed thematic areas (though they could also search as well) to locate a related informational object and could immediately see or hear the information while discovering the type of information (e.g., text, audio, and images), title, author, description, date and “theme” for the information. . This

condition was intended to present the same resources as in the control condition but in a cognitively flexible format that was hypothesized to enhance the ability of learners to synthesize and integrate new knowledge.

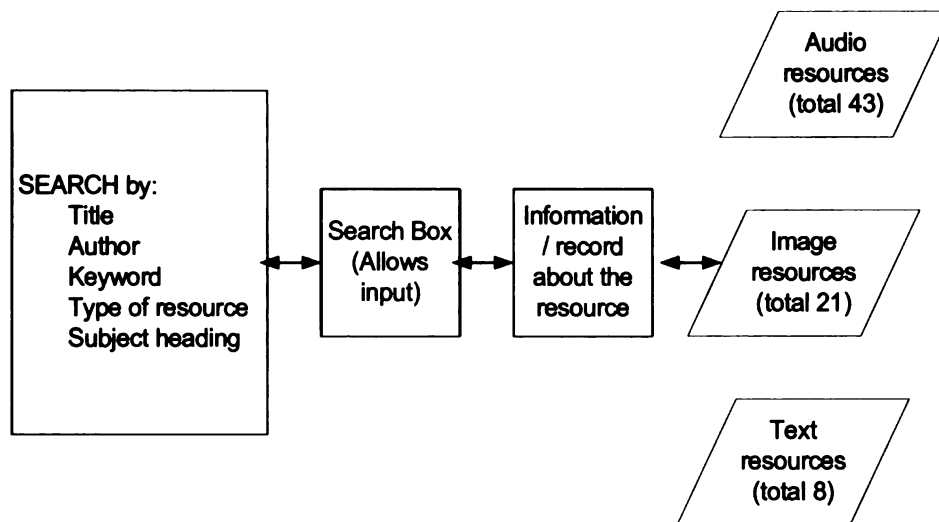


Figure 1-3. Control Condition Site Map

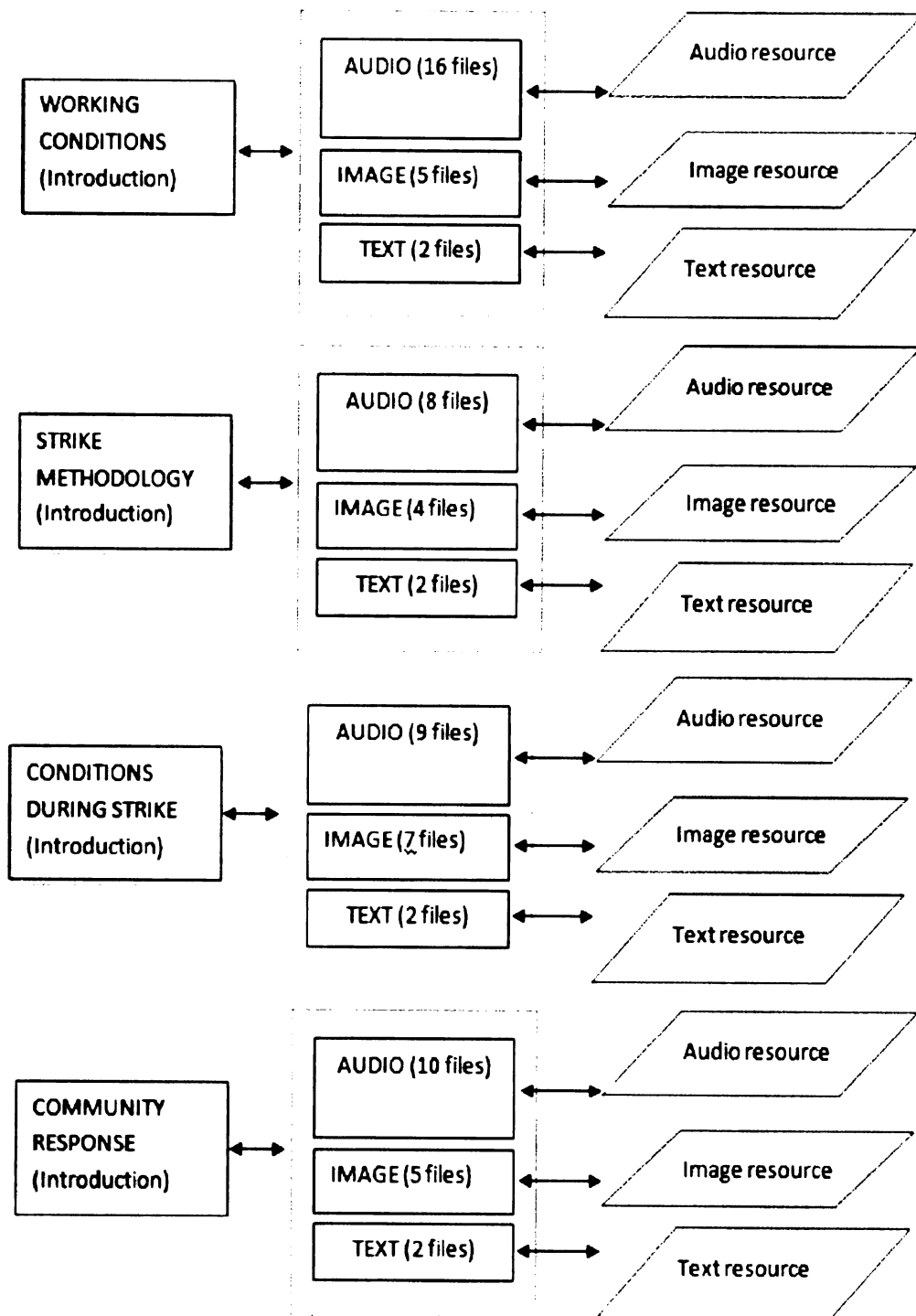


Figure 1-4. Experimental Condition Site Map

Through context-rich presentations, it was expected that learners would cognitively benefit by the creation of "intellectual erector sets' (which permit) open-ended exploration in the context of some flexible background structures, (and) aspire to the goals of making knowledge a manipulatable, 'three-dimensional' entity for the learner..." (Spiro et al., 1992, p.125). For example, like physical erector sets knowledge is often composed of different segments (e.g., brief facts, definitions, time frames, explanations, etc.) that can be recombined to form new knowledge "structures." It was also expected that learners in the linear condition would have higher test results for factual knowledge items (Jacobson & Spiro, 1995). Further, it was believed that complementary sources of visual and verbal information co-presented would aid learning (Clark & Paivio, 1991). Therefore it was hypothesized that the more rigidly structured knowledge in the control condition would result in comparative differences when compared with the more accessible and flexibly-structured knowledge in the experimental condition. Specifically, it was hypothesized that:

1. participants in the linear, control condition would achieve higher scores than participants in the context-rich, experimental condition on the test for factual recall;
2. participants in the context-rich, experimental condition would achieve higher scores on the test for synthesis than participants in the linear, control condition;

3. participants in the context-rich, experimental condition would achieve higher scores on the test for integration than participants in the linear, control condition;
4. participants in the context-rich, experimental condition would achieve higher scores on the test for retention than participants in the linear, control condition, and
5. participants in the context-rich, experimental condition would rate their system more favorably than participants in the linear, control condition.

Limitations

The major limitations of this study are related to statistical sampling issues used in this experimental design. In sum, it is not possible to generalize the results of this study beyond the experiment's sample of participants and their treatments as the participants were not randomly sampled. In addition, the design of hypermedia systems is endlessly varied and the technology and standards used to build them changes. This means that it is difficult to compare the design and subsequent results of different systems. Lastly this study followed learning (up to two hours) and testing procedures (immediate and one week following exposure; facts, integration, synthesis and retention), which cannot be directly applied to other research.

CHAPTER 2. LITERATURE REVIEW

There have been no empirical studies to date that have examined how online library catalog interfaces differ from non-linear, and more complex, context-rich hypermedia interfaces. The lack of research is striking, as despite the lack of an empirical foundation, there is a strong belief that alternative digital library interfaces (ADLIs) may be a superior form for accessing and understanding information in contrast to library catalogs. Examining the cognitive effects of ADLIs and online catalog interfaces is important for three reasons. First, from a conceptual perspective, ADLIs and their role as possible learning environments warrant better understanding. That is, ADLIs and online catalogs are often considered integral to the research process, for the identification and collection of materials on a given topic. However, what is not clear is how they affect the researcher's thinking and understanding of their topic. Second, from a resource and policy perspective, ADLIs are a significant area of research and funding, so understanding their impact on cognition is important. The National Science Foundation's ⁹ *Digital Libraries 2* initiative from 1999 through 2005 funded \$49,225,417 worth of research and development. Other national funding entities that award monies include the Library of Congress, National Library of Medicine, National Endowment for the Humanities, National Aeronautics and Space Administration and the Institute of Museum and Library Services. Third, from both a conceptual and resource perspective, ADLIs are often developed in tandem with online library catalogs. This is not to say that either effort is necessarily

⁹ <http://www.dli2.nsf.gov/projects.html>

duplicative or superior, but the fact is that the Library of Congress,¹⁰ for example, estimated that the cost per volume (e.g., book) for original cataloging in the fiscal year 2001 through 2002 was \$94.58. Therefore, the total cost of cataloging for the Library in that fiscal year was \$29,342,026. This cost per volume and the known difficulties patrons have in accessing library materials (Bates, 2003; Sandberg-Fox, 2001; Borgman, 1996; Drabenstott, 1991) have resulted in diminished support for the cataloging process itself. At the same time however, funding for ADLIs has increased. It is therefore important to understand within an empirical framework the comparative constraints and affordances of online library catalog and ADLIs.

Relevant Empirical Literature

While there have been no empirical studies comparing online library catalog interfaces and digital library interfaces from a cognitive processing perspective, there have been many studies exploring the constraints and affordances of hypertext and later hypermedia designs and their effects on cognitive processing. Following Dillon and Gabbard's (1998), review entitled "Hypermedia as an Educational Technology: A Review of the Quantitative Research Literature on Learner Comprehension, Control and Style," an initial search in the Educational Resources Information Center (ERIC)¹¹ database and the PsychINFO¹² database for the period 1996 – 2004 was undertaken for relevant empirical research (Wells, 2005). This initial search overlapped with

¹⁰ <http://www.loc.gov/faq/catfaq.html#11>

¹¹ ERIC was searched using the terms ((mj: hypermedia and (mj: learning or (mj: instructional w effectiveness) or (mj: instructional w design)))) and limited to (yr: 1996-2004)

¹² PsychINFO¹² was searched using the keyword query (hypertext or hypermedia) and (cognit* or learning or study) and limited to (PY:PY = 1996 – 2004)

Dillon and Gabbard's comprehensive review but assured that literature that was published or indexed in 1996 was included. Three years later, this initial search was updated and two additional databases were searched. Specifically, Education Abstracts was searched for the period 2004-2007 using the terms "su: hypermedia or su: multimedia or su: hypertext and su: learning or su: cognition" was undertaken for relevant and up-to-date empirical research. PsycINFO was also searched for the period 2004 – 2007 using the terms "kw: hypertext or kw: hypermedia or kw: multimedia and kw: cognition or kw: learning or kw: study." Lastly, an updated review from Dillon and Jobst (2005), entitled "Multimedia Learning with Hypermedia" was checked in order to ensure that pertinent studies were not missed. From these two searches, sixty empirical and relevant studies were located. Articles with a poor experimental design (e.g., did not control for differences in reading comprehension, had minimal instructional time, a minimal number of participants, etc.) and those whose major results focused on attitudinal or efficiency outcomes were excluded.

The research from 1997 – 2007 can be grouped into four broad themes

including:

1. individual differences,
2. structural differences,
3. efficiencies and
4. satisfaction, motivation and attitudes.

Studies on the first theme, individual differences, focus on the identification of personal characteristics that influence learning. Individual differences included background knowledge, academic discipline or major, cognitive styles (e.g., field-dependence/independence, holist/serialist biases, verbaliser/imager, etc.), computer experience, age, gender, motivation, speed and attitude.

Studies that fall under the second theme, structural differences, focus on the system structure and interface design. Studies on structural differences can be grouped in five main areas: presence or absence of an explicit content structure, division of content, organization of content, navigational styles and freedom of movement. For example, Brinkerhoff, Klein, & Koroghlanian (2001) studied the effects of structured and unstructured content and the presence or absence of overviews (i.e. summative document with an organizational map). Dee-Lucas & Larkin (1999) studied more and less segmented text, which refers to the relative division of content into separate topics. Waniek, Brunstein, Naumann & Krems, (2003) studied the effects of hierarchical vs. linear vs. chronological displays of content. Ford & Chen (2000) studied navigational styles including patterns and depth of individual movement. Zumbach, Reimann, & Koch (2001) studied the relative level of freedom individuals had in determining how to view content.

Research related to the third theme, efficiencies, typically sought to find a connection between the time individuals spent reading and/or learning and their subsequent achievement. Examples include the time it takes for an individual to learn and subsequently answer questions (i.e., temporal efficiencies) and the amount of mental effort that is required by an individual to learn (i.e., cognitive

efficiencies). The goal is typically to reduce either the temporal and/or cognitive requirements while raising achievement. For example, Lee and Tedder (2004) studied the effects of three types of text: linear or scrolling hypertext, structured hypertext and an unstructured hypertext. When reading time was not controlled, individuals in the linear condition scored the highest. However, when reading time was controlled there were no significant differences in recall between the conditions.

Studies that focus on the fourth theme gathered and/or assessed participants' satisfaction, motivation and attitudes toward system design and/or learning. For example, Triantafyllou, Pomportsis, Demetriadis & Georgiadou (2004) "...included items relating to the completeness and ease of use of the system as well as items on the subject's satisfaction and willingness to use the system" (p. 102). Motivation was assessed pre-test or post-test. In pre-tests motivation was typically assessed as an individual characteristic that affected achievement (Liu, 2006) whereas in post-tests, motivation was typically seen as a function of the hypermedia and/or system design (i.e., does the system motivate individuals) (e.g., Liaw, Huang & Chen, 2007). Attitude was often defined in terms of an individual's feelings toward a subject or the value they placed on learning (Liu, 2006).

Individual Differences

Schnotz (1999) (as quoted in Tardieu & Gyselinck, 2003) notes "the temptation is strong to simply assume that using multiple forms of displaying information...results generally in better learning" and this assumption is evident in

the empirical research (p. 3). The assumption concerning multiple displays of information is that increased learning can occur given an optimum design configuration that matches a learner's need and characteristics. For example, Ford and Chen (2000) designed a hypermedia system that contained seven different navigational options including a topic map, keyword index, top-level menu, section buttons and subject categories in order to "...explore the effects of individual differences on learners' navigation patterns (with) resultant learning outcomes" (p. 282). Though the authors report on several findings, they conclude that there is no difference in navigational strategy and outcome. Mitchell, Chen, & Macredie (2005) examined differences in prior domain knowledge or expertise and an individual's comfort with linear and non-linear pathways. They find significant differences in learning for those individuals who scored the lowest in pretests for subject knowledge but also find that these individuals were more disoriented within the non-linear system. Liu and colleagues (Liu, Bera, Corliss, Svinicki, & Beth, 2004) studied different patterns of cognitive processing, including information processing, which was defined as "problem solving," and metacognition, which was defined as "thoughtfulness," and an individual's use of cognitive tools to support learning. However, while they find significant difference from pre-test to post-test scores, there were no significant differences between the two groups on the post-test scores. Triantafyllou, Pomportsis, Demetriadis, & Georgiadou (2004) report on an experiment to test cognitive styles or how individuals process information (e.g., field independence versus field dependence) in one of two conditions, an adaptive hypermedia and a traditional hypermedia

environment. Though they find a significant difference between environments, they do not find a correlative difference for cognitive style.

Structural Differences

Many researchers focused on structural elements: hypertext/hypermedia, linear displays, or the absence of overviews and their subsequent effects (Huk & Steinke, 2007; Bernard, Hull, & Chaparro, 2005; Calandra & Barron, 2005; Eveland, Marton, & Seo, 2004; Lee & Tedder, 2003; Brinkerhoff, Klein & Koroghlian, 2001; Ford & Chen, 2000; Dee-Lucas & Larkin, 1999; Hofman & van Oostendorp, 1999). For example, Huk & Steinke (2007) study the effects of "...visualization strategies for structuring non-hierarchical learning tasks..." by contrasting close-up views of cells (p. 1089). One display included only an enlarged image of the relevant cell part and the other displayed the relevant cell part in the context of the whole cell. Both systems provided equivalent amounts of information. The part within the whole condition produced higher narrative recall, while an effect for transfer of knowledge was dependent on having a high spatial ability and being in the part within the whole condition. Eveland, Marton, & Seo (2004) investigated the relationship between hypermedia systems design and "...the making of mental connections among pieces of new information as well as between new and old information in memory" (p. 89). Using content from the *New York Times*, two conditions were created. For the first, the indexed condition, only three stories were contained on the home page and the rest could only be accessed through a categorized list (e.g., national, international, politics, etc.), which was also on the home page. For the second, the linked condition, index

links were available on the home page and between the stories. Though it was hypothesized that the linked condition with its interconnected knowledge structure would support the development of greater knowledge, no significant effect was found. The researchers did however find a significant difference in factual recall for those using the indexed condition, which was composed of topic lists, versus those using the linked condition. Hofman and van Oostendorp (1999), after studying the effects of structural overviews versus topic-lists, report that participants with low prior knowledge may have difficulty developing a situational or overall mental model with overviews. Dee-Lucas and Larkin (1999) examined what participants in more and less text segmented groups (i.e., text containing more or less nodes) could recall of the content. They found that learners using the more-segmented hypertext recalled a narrower range of content while learners using the less-segmented text recalled a broader range of content. They conclude that different learning goals such as depth or breadth of content coverage should determine the segmentation of hypertext content.

A second variant of structural differences includes advanced organizers, which are designed to scaffold learning by making hypermedia structures explicit. Calandra & Barron (2005) investigated the effects of an advanced organizer with text, an advanced organizer with text and graphics and a control condition with no advanced organizer. The researchers report that no significant difference in a test for knowledge between the experimental conditions and the control conditions but note that future tests might explore problem solving and not just factual understanding.

A third variant of structural differences includes mixed media (i.e., the use of multiple forms of media). Zahn, Barquero, & Schwan (2004) studied the effects of more and less sequential and clustered links with video and text using 4 experimental groups and 1 control group. They conclude that knowledge acquisition was not affected by condition rather, "...subjects could learn comparably well with all four hypervideo designs and with the text and video materials presented without hyperlinks" (p. 284).

Efficiencies

It is interesting to note the importance of efficiencies in the literature. For example, Nadolski, Kirschner, & van Merriënboer (2006) developed two systems with the same amount and type of content. The first was highly segmented while the second was lightly segmented. They then studied the relationship of task to efficiency, which included performance, mental effort, time on task and motivation, and found that participants in the lightly segmented conditions were more efficient than participants in the highly segmented conditions. That is, participants in the lightly segmented conditions outperformed the participants in the highly segmented conditions on the learning tasks and required less time to complete the tasks. There was no difference though between conditions for mental effort or motivation. However, most of the efficiency measures in the literature focus on temporal dimensions. In 2003 Waniek, Brunstein, Naumann, and Krems tested three different text conditions (i.e., hypertexts) and find no significant difference in total time spent in each condition though they had hypothesized otherwise. Further, they conclude "participants under different text conditions did not differ in

their factual knowledge of the content though, again, they had hypothesized otherwise" (p. 109). Likewise, Dee-Lucas and Larkin (1999) also studied lightly and highly segmented text find no differences between the amount of time participants spent studying the given content in each condition and in tests for content knowledge. Oostendorp and Nimwegen's (1998) come to a similar conclusion after studying how long it took participants to locate information while using longer and shorter texts. They found no difference between search times, the percentage of search tasks successfully completed and the amount of information participants could recall. They do find however, that when information is outside the screen border and requires two (or more) clicks to access, performance, which was defined as an individual's ability to locate specific information, was poor.

Satisfaction, Attitude and Motivation

Additional individual differences including satisfaction, attitude and motivation are prevalent in much of the research. Specifically, a positive correlation is reported between hypermedia and the participants' overall feeling about the method of instruction (Su & Klein, 2006; Gauss & Urbas, 2003; Waniek, et al. 2003; Brinkerhoff et al. 2001; Dee-Lucas & Larkin, 1999; Oostendorp & Nimwegen, 1998). Su & Klein (2006) studied the effects of embedded links, content list and concept map on attitudes. They found that the content list condition was rated most positively, followed by the concept map condition with the embedded links condition coming in last. Brinkerhoff et al. (2001) explore the effect of content summaries and attitudes on learning. They find that the presence

of a summary had a significant and positive effect on attitude; however, the presence of a summary did not affect achievement. They conclude that their learning environment may have been well organized to begin with, thereby rendering their summaries useless. Gauss and Urbas (2003) explore individual differences, navigation and learning outcomes. They find that intrinsic motivation had a positive effect on learning regardless of prior learning and, in general, that intrinsic motivation, attitude toward computer-based learning and computer experience were significantly interrelated. They conclude that there is a need for stronger motivational design, which uses motivational concepts and theories to influence individual learning (Song & Keller, 2001).

Summary

This literature informs the development of learning environments, including ADLIs. In addition to understanding theory, knowing what approaches have been taken to the organization, design, assessment, etc. of learning environments has implications for development of ADLIs. For example, it is important to know that while many types of individual differences such as learning styles, experience, discipline etc. have been theorized as significant in the development of hypertext/hypermedia systems, low prior knowledge may be most significant. Specifically, individuals with low prior knowledge seem to have more difficulty than those with high prior knowledge in comprehending new material (Zumbach, 2006). As Dillon and Jobst (Dillon & Jobst, 2005) note "there is an unfortunate irony to this as hypermedia has long been advocated as a way of 'leveling the playing field'" (p. 257). Like individual differences, structural

Differences are another area that has generated a great deal of research-related interest. Differences have included hierarchical versus linear displays, the presence or absence of content maps or summative and findings, etc. Structural differences can aid information retrieval and, consequently, learning. Although no specific structural difference is most clearly linked to learning outcomes, content lists and hierarchies seem to be preferred over other forms such as maps and embedded links (see for example, Su & Klein, 2006; Bernard, Hull, & Chaparro, 2005; Eveland, Marton, & Seo, 2004). Assessment of attitudes is likewise significant. At a fundamental level, hypermedia learning environments must be usable and create favorable impressions in their users. Findings in this research have direct implications in the design of ADLs.

Cognitive Flexibility Theory

There is however, one additional study, which was previously mentioned, that is particularly significant. In 1995, Jacobson & Spiro exposed participants to two different hypertext systems with structural differences. One of the learning environments, the experimental condition, incorporated the CFT principles of multiplicity, interconnectedness and adaptive flexibility, which acknowledge and manage complexity (See Table 2-1. Cognitive Flexibility Principles), while the other, the control condition, was a more rigidly-structured hypertext. In this study, participants in the control condition completed the study stage in significantly shorter periods of time and acquired more factual knowledge than did those in the experimental condition. However, participants in the experimental condition

achieved higher problem-solving scores than participants in the control condition (Jacobson & Spiro, 1995).

Cognitive Flexibility Theory (CFT) suggests seven principles that should guide instruction in ill-structured and complex domains (See Table 2-1. Cognitive Flexibility Principles). These principles are: Multiplicity, Complexity, Context-dependency, Interconnectedness, Inexhaustibility of Understanding, "Openness" in Conceptual Structures and Adaptive Flexibility (Mishra et al., 1996; Spiro et al., 2004; Spiro, Coulson, Feltovich & Anderson, 1988; Spiro et al., 1987).

Table 2-1. Cognitive Flexibility Principles

-
- 1. Multiplicity/ Multiple knowledge representations** – “Information that has to be used in many ways has to be represented in many ways” (Spiro et al., 1987, pp. 187-188) with the effect that students can develop “open knowledge structures” that permit them to selectively and creatively apply knowledge rather than develop rigid and incomplete understanding of the domain.
 - 2. Complexity** – “The introduction of complexity at the initial stages of the instructional process (albeit in manageable chunks) guards students from being seduced by or seeking inappropriately simplistic interpretations and understandings in complex and ill-structured knowledge domains.” (Mishra et al., 1996, p.7).
 - 3. Context-dependency** – “Facts do not remain self-evident, isolated bits of information but, rather, are “constructed” by their perceived relationship to other facts and by their usefulness in understanding cases” (Mishra et al., 1996, p.8).
 - 4. Interconnectedness** – “Conceptual and case knowledge can not be ‘boxed’ into separate mental compartments [but rather requires juxtaposition in order to support] the goal of widely applicable or transferable knowledge” (Spiro et al., 2004, p. 6-7).
 - 5. Inexhaustibility of Understanding** - “...CFT promotes the “revisiting” of cases and thematic explorations...(and tries to promote) the excitement of seeing the same thing with a new and different set of “lenses”...” (Mishra et al., 1996, p.10).
-

Table 2.1 (cont'd).

6. "Openness" in Conceptual Structures - "...what is provided are open structures to help one start in one's construction of new knowledge, rather than closed structures that restrict constructive activity" (Mishra et al., 1996, p.11).

7. Adaptive Flexibility - "The main aim of CFT hypertexts is to help students acquire flexible cognitive skills that can take multiple, interrelated concepts and apply them to new, diverse and largely unexpected circumstances..." (Mishra et al., 1996, p.11).

The present research seeks to build on Jacobson & Spiro's (1995) study. Specifically, the present study replicates in many ways the earlier study. This is important as Jacobson & Spiro (1995) succeeded in measuring significant differences between conditions on the acquisition of factual knowledge and were able to demonstrate that structural differences could positively affect problem-solving in contrast to much of the literature. Like Jacobson & Spiro (1995), both studies concern historical events, the impact of technology on 20th century society and culture and the Flint-sit-down strike. Both pre-tested undergraduate participants from large mid-western universities for domain knowledge and verbal comprehension in order to control for pre-existing knowledge, reading skill and general intelligence. This is in contrast to the previously mentioned studies, which tended to pre-test only for domain knowledge. Both studies use problem-solving measures in addition to factual tests and test retention. As R.E. Mayer (as quoted in Calandra & Barron, 2005) notes problem solving, which is referenced as

“meaningful learning” (p. 20) as well as retention need to be measured in hypermedia research, yet few studies have done so. Therefore, this study by building on successful research is an important addition to the literature.

Why is this research important?

This study is important for other reasons as well. In 2005, Dillon & Jobst make two relevant points during their review of the hypermedia literature. First, “it seems that structure will affect learning by influencing how well or how fast a learner can move through a hypermedia document” (p. 257). Key to this point is the idea that the structure of content helps to determine whether and how ably individuals can navigate content. It makes sense that the more ably an individual can deliberately navigate content, the greater their ability to develop some cogent understanding of the material. Therefore, this study, adds to the research on an important issue in hypermedia research – structural differences. Second, as is evident in this review but also noted by Dillon & Jobst (2005), many researchers have focused on studying text and text and images rather than including other media such as audio as well. Therefore, this study is relatively unique in that it includes text, images and audio and its conclusions concerning the effects of multimedia help to fill a gap in the literature.

An additional gap this research fills is the need for studies whose conditions have undergone usability testing prior to testing. Of the previously mentioned studies involving structural differences, only Calandra & Barron (2005) tested their systems and then mitigated any usability problems. This is significant because there are no standard hypermedia designs, unlike, for example, printed

books. Systems can be almost endlessly varied. For example, the organization of materials, content, tasks, issues of accessibility, labels, feedback, response times, labeling/language, colors, fonts, etc. can vary widely. What this means is that researchers may be making incorrect assumptions regarding how usable their systems are including how functional the system is and whether it meets its purpose and this may confound subsequent results. The present study, by having attempted to control for extenuating factors may afford clearer results.

Lastly, also different from many of the previously mentioned studies is this studies' focus on the effects of different displays of hypermedia information, a simulated online library catalog and digital library, a heretofore unexplored area of study. While other researchers have studied the presentation of segmented content, similar to that of the control and experimental condition, individual differences in learning and/or how people search for and locate information in hypermedia environments (e.g., Bera & Liu, 2006; Lee, 2005; Lee & Tedder, 2004; Schwartz, Andersen, Hong, Howard & McGee, 2004; Oostendorp and Nimwegen, 1998) no one has contrasted these two types of learning environments.

In summary, this proposal seeks to contribute cognitively-based measures on the effects of different displays of hypermedia information. It builds on the existing literature in the field by focusing on structural differences in general and on research from Jacobson and Spiro (1995) in specific. It adds to the hypermedia literature by its use of audio, text and image resources in contrast to much of the literature. Because the learning environments were subjected to heuristic testing

and the subsequent mitigation of usability issues, the results are not confounded by any difficulties associated with how the systems were implemented. This research also makes a unique contribution because of its focus on the comparative effects of a simulated online library catalog interface and a digital library interface. Lastly, this study will add to the literature surrounding CFT, which is significant because much of the literature in the area of information science and educational technology is atheoretical (Pettigrew et al. 2001; Yang, 2001; Spiro & Jehng, 1990).

CHAPTER 3. RESEARCH METHODS

This chapter is divided into five sections. The first, Testing and Learning Stages explains the multi-stage process to which participants were exposed .The second, Background and Materials itemizes and explains the procedures that occurred prior to data collection and the data collection environment itself. The third, Testing Procedures, covers the actual testing phases and specifics on each of the measures given or on the participant information that was gathered. The fourth, Scoring Procedures, provides specifics on how all measures were assessed. The fifth, Statistical Tests, summarizes the tests used in the following chapter on Statistical Analysis.

Testing and Learning Stages

Participants were exposed to one of two systems in a multi-stage process (Figure 3-1. Testing flowchart), which required two hours to complete in a campus lab. In part one, the *initial testing and learning stage*, participants were randomly assigned to the linear, control condition or the context-rich, experimental condition. After arriving at the lab, they were pretested for prior knowledge of the Flint Sit-Down Strike. They then received a complete list of all the resources available (Appendix K: Resources) and were told they could use this list to keep count of those resources they listened to and read. Participants were instructed to listen to twenty audio recordings and read four text files in one of two conditions. These requirements were intended to ensure that all participants, regardless of condition, were exposed to the same number and

format of resources. That is, participants in neither condition were advantaged or disadvantaged by knowing or not knowing how to manipulate the interface. There was no minimum or maximum amount of time they were required to spend in their condition.

In part two, the *factual recall, synthesis and integration test stage*, participants were asked to recall bibliographic, factual details for six resources and prepare two 250-word essays. The first essay tested for knowledge explicitly gained and integrated within and across two or more themed areas. The second essay tested for knowledge gained and synthesized from the first of the themed areas. Successful responses to both sets of essays required knowledge of the resources in different theme areas. These essays were intended to highlight differences in learning based on the contrasting conditions. The order of the tests (i.e., testing for integration within and across two or more themed areas before testing for synthesis within a theme) was intended to avoid allowing participants a period of reflection and action that might change the participants' ability to integrate content. Lastly, the participants completed an attitude survey, took a vocabulary test and answered some background questions.

In part three, the *retention test stage*, participants were asked to answer ten questions and complete one short answer question. The retention survey required approximately ten minutes to complete and was designed to test their factual recall and its synthesis of material concerning the Flint Sit-Down Strike one week after their exposure to either condition. A link to the survey was emailed to

participants who completed it wherever they wished (e.g., home, library, classroom, etc.) within a forty-eight hour period.

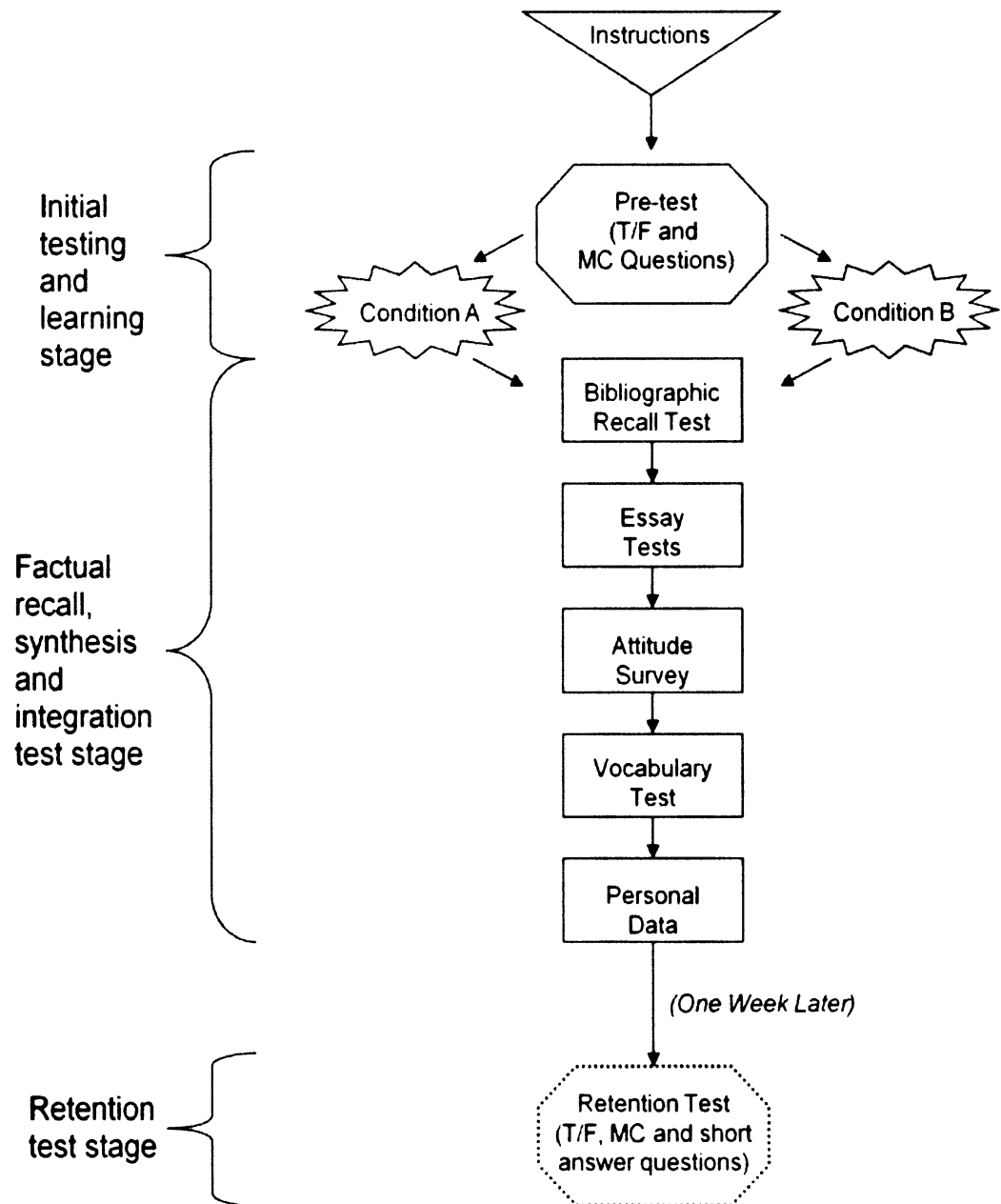


Figure 3-1. Testing flowchart

Background and Materials

The Flint Sit-Down Strike of 1936-1937 was a strike of national importance as it established the United Auto Workers (UAW) as the bargaining representative for workers at the world's largest corporation, General Motors. Flint was the first site in a series of sit-down strikes that ultimately spread to include "approximately 135,000 men from plants in 35 cities in 14 states" (*The Detroit News*, n.d.). After 44 days and the intervention of President Roosevelt, the strike was settled by management and labor.

This study used different approaches to structuring hypermedia information on the Flint Sit-Down Strike. The hypermedia information was comprised of oral histories in audio form, short and long texts, and images from *The Detroit News*. Both hypermedia learning systems used the same objects (texts, audio, and images). The majority of the text and all of the images were contemporaneous with the strike. All of the text, images and audio were selected by Drs. Michael Van Dyke, Visiting Assistant Professor, Department of American Thought and Language, and David Bailey, Associate Professor, History Department at Michigan State University. Drs. Van Dyke and Bailey also wrote the non-contemporaneous text. The audio interviews were conducted from the early 1970s until 1980 and are first-person accounts by people who participated in, assisted with or fought against the strike itself.

The content covers four aspects of the Flint Sit-Down Strikes: “working conditions,” “strike methodology,” “conditions during the strike” and “community response.” “Working conditions” includes issues such as wages, assembly line speed, water and food breaks, job security and the extreme heat within the plants. “Strike methodology” includes issues such as secrecy, timing, the theory behind this new type of strike, the role of women and the company’s response. “Conditions during the strike” includes issues such as heating the plant and the organization of food and communications between the strikers inside the plant and people outside the plant. “Community response” includes issues such as the reaction to union organizers and the strike itself from pro- and anti-union people in the Flint area.

In summary, a “between subjects” manipulation with 33 participants was followed in a campus computing lab. The studies’ independent variables were knowledge of labor history, condition, verbal comprehension, gender, familiarity with library catalogs, the amount of materials checked out from a library in the most recently completed semester and the amount of time spent searching the Internet and background pretest score. The dependent variables were attitude, achievement (recall, synthesis, integration and retention), time spent exploring the system and time spent completing the achievement tests (see Table 3-1. Summary of Variables). These are explained below.

Table 3-1. Summary of Variables

Type of Variable	Specific Variable
Independent variables	<ol style="list-style-type: none"> 1. Gender 2. Condition 3. Verbal comprehension 4. Background pretest 5. Familiarity with library catalogs 6. The amount of materials checked out from a library in the most recently completed semester 7. Familiarity with searching for information
Dependent variables	<ol style="list-style-type: none"> 1. Attitude 2. Achievement (recall, synthesis, integration and retention) 3. Time spent exploring the system 4. Time spent completing the achievement tests

Participants were asked to supply limited personal information and were tested for background knowledge prior to the intervention. Participants were *not* given specific instructions on what to concentrate on, though they *were* informed that they would be asked to answer a series of factual test questions and two essay questions based on the information to which they were exposed. Further they were instructed that they would be asked to complete a follow-up survey one week following exposure. They were told to "...try to put (all the information) together as best you can..." and "...to learn as much as you can about the Flint strike..." Following the intervention, participants' immediate achievement was measured via the completion of a test for recall and their responses to two essays (See Appendix C: Evaluation Measures). Lastly, they completed an attitude survey (See Appendix D: Attitude Survey) and vocabulary test (Ekstrom, French, & Harman, 1976) and answered some basic background questions (See Appendix F: Personal Data). One week following the intervention, participants' achievement was again measured via the completion of a test for retention. (See **Error! eference source not found.** for a graphical flowchart of the testing.)

Sample

The sample was drawn from self-declared undergraduate history majors at Michigan State University, a large public land grant university located in the Midwest. As a whole, Michigan State University accepts approximately 70% of those applying with over 80% coming from within the state. Out of an incoming freshman class of over 6,000 students, approximately 600 students are African-

American, 300 are Asian/Pacific Islander, 150 are Chicano/Hispanic and 50 are American Indian (Cotter, 2006). In the Fall of 2006, Cotter stated that the

...average ACT and SAT scores for incoming freshmen met or exceeded both state and national averages. For the ACT, the mean score of the incoming class was 24.6, compared to an average of 21.5 for Michigan and 21.1 nationally. For the SAT, the mean score of the incoming class was 1151 — above the U.S. average of 1021 and equaling the average of all Michigan students. (p. 2)

Self-declared history majors at Michigan State University were recruited via flyers posted within the History department, a History department mailing list for all declared (primary or secondary) history majors and through both History and Teacher Education faculty announcements (see Appendix G: Recruitment Form). Participants were paid \$20.00 or \$25.00¹³ to complete survey 1 (i.e., parts one and two, the *initial testing and learning and test stages*) and \$5.00 to complete survey 2 (i.e., part three, the *retention test stage*). Participants were randomly assigned a condition using a "between subjects" manipulation. That is, while participants were given a choice of testing dates and times (Wednesday, 1-3 or 4-6, or Friday, 10-noon, 1-3, or 3:30-5:30) over the course of four weeks (April 4th - April 27th), the factors that determined which condition a person was assigned to were their own personal scheduling needs and restrictions on group

¹³ The amount of compensation was increased in order to recruit additional participants. Twenty-seven initial participants received \$20.00 and six secondary participants received \$25.00.

size. Sixteen males and seventeen females were assigned to one of the two conditions, the simulated online library catalog or the simulated digital library, with each participant using and subsequently being tested on one of the conditions.

Table 3-2. Summary of Participants				
	Online library catalog		Digital library	
Participants	Male	Female	Male	Female
	8	9	8	8
Total participants (n=33)	17		16	

The number of participants was supposed to be restricted to no more than 6 and no fewer than 2 per testing session to ensure that the author/investigator was able to respond to participant questions. However, in practice sessions were conducted with 1 to 7 participants¹⁴. Sessions of up to two hours were conducted until data from 33 participants were collected.

¹⁴ Though the sessions were limited to no more than 6 and no fewer than 2, in fact, during one testing session with 6 participants, one person who was not scheduled to be tested, showed up and was included in the session. In addition, one other session had only one participant when another confirmed participant did not show up.

Heuristic Testing

Prior to testing participants it was important to make sure that each system functioned properly and that the design of each while remaining realistic wasn't problematic. That is, prior to testing assumptions regarding recall, synthesis, integration and retention, both systems (i.e., the control and the experimental condition) needed to be evaluated to determine if the underlying system structure (including fields, URLs/linking, search engine, system speed, etc.) and interface (including typography, colors, layout and design, help, navigation, forms, etc.) at best, facilitated use and, at worst, did not hinder use. That is, both the online library catalog and hypermedia learning environment's overall designs needed to be at least neutral for the purpose of understanding subsequent learning.

Therefore, heuristic testing was used to determine whether the two systems satisfied the needs and expectations of evaluator-participants (hereafter referred to as "evaluators"). As is common in usability testing, this meant that any and all issues identified as major or catastrophic problems by the evaluators had to be mitigated. For example, if one or more evaluators judged a label as unclear or color coding as inappropriate and rated the issue as major or catastrophic, as was the case, then the specific issue had to be mitigated by changing the name of the label and the color-coding.

In keeping with Nielsen (n.d.) 4 evaluators without domain expertise but all of whom had been working in the field of usability for a minimum of three years ($\mu = 7.5$ years) examined the system and then completed the checklist the fall of 2004. The evaluators chose whether to conduct the testing in their own office or

the author/observer's office. Each evaluator worked independently but with the author/observe present. They then completed the checklist (See Appendix A: Heuristic Evaluation). In addition, each session was audio recorded. Their ages ranged from 28 to 45 years ($\mu = 37$ years). Two held doctorates, 1 had submitted her dissertation and 1 was working toward a Bachelor of Arts.

Heuristic evaluations typically consist of "...four phases: a pre-evaluation training session, the actual evaluation, a debriefing session to discuss the outcome of the evaluation, and a severity rating phase during which the evaluators assess the severity of the usability problems that had been found in the evaluation session" (Nielsen, 1994). However, two phases were merged by the evaluators, the evaluation and severity rating phase, which meant that instead of one phase sequentially following the other, the evaluators examined and assigned ratings for any given item prior to proceeding to the next item (See Table 3-3. Phases in the Heuristic Evaluation).

Table 3-3. Phases in the Heuristic Evaluation

Pre-Evaluation Training Session

Heuristic Evaluation and Severity Rating

Debriefing Session

Each evaluator inspected both systems (i.e., the control and the experimental condition). All were asked to explore the two working systems independently and the heuristic instrument (See Appendix A: Heuristic Evaluation) prior to beginning (Nielsen, n.d.). Each evaluator had one to two hours to complete the written

evaluation. All comments were recorded following a talk aloud protocol (Nielsen, 1994; Ericsson & Simon, 1980). Table 3-4. Summary of the Heuristic Evaluation – components and procedure” provides an outline of the process and procedure followed.

Table 3-4. Summary of the Heuristic Evaluation – components and procedure

Component	Procedure
Methodology	Heuristic evaluation
Objective	Determine compliance with usability principles
Scenarios	Independent
Timing	Following the development of a working system
Population	Non-domain experts
Number in population	3-5 evaluators
Training and background of population	Non-domain experts who, as professionals, develop and/or evaluate systems. Some heuristic training was given.
Test Presentation	Working system
Setting	Individual evaluators with author/observer present.
Duration	1-2 hours per evaluator.
Method of response	Written reports with verbalizations included.
Author/Observer	Present
Data	Quantitative via written report and severity rating. Qualitative via audio recordings.

A modified version of the instrument “Heuristic Evaluation: A System Checklist” (Pierotti, 1995) was used (Appendix A: Heuristic Evaluation - A System Checklist). The categories of potential issues included:

1. **Visibility of System Status** (i.e., degree to which the user can follow what is going on in the system),
2. **Match between System and the Real World** (e.g., the user's language rather than systems language should be used),
3. **User Control and Freedom** (i.e., users should maintain a degree of freedom to move around the system),
4. **Consistency and Standards** (i.e., different words, situations, or actions should have comparable meanings),
5. **Help Users Recognize, Diagnose, and Recover From Errors** (i.e., errors should use plain language),
6. **Error Prevention** (i.e., are images, text and audio easy to access),
7. **Recognition Rather Than Recall** (i.e., user should not have to remember information),
8. **Flexibility and Minimalist Design** (i.e., is there redundant and/or distracting information used), and
9. **Aesthetics** (e.g., only pertinent information should be displayed).

The data collected were quantitative and qualitative. Data were examined to ensure that there was consistency in reporting (i.e. that there were no discrepancies between something the evaluator mentioned during testing but did not physically note) in order to ensure the most comprehensive redesign.

Some discretion is required to understand these results as an item need only be rated catastrophic by one participant and not by any other participant to be singled out as a usability problem. Still, following the heuristic testing, all

major and catastrophic issues were compiled into a written report and then mitigated whenever it was possible to do so. (This is explained more fully below).

In summary, the control condition, the online library catalog had one item that was rated as a catastrophic usability issue: "images, text and audio files/sounds are not easy to access." The context-rich learning environment, had nine items that were rated as catastrophic usability issues:

1. "continuity of thinking is required to remember information through several screens,"
2. "high levels of concentration necessary for remembering information from screen to screen,"
3. "menu choices are not ordered in the most logical way, given the item names, and the task variables,"
4. "there is a natural sequence to menu choices, but it has not been used,"
5. "related and interdependent information does not appear on the same screen,"
6. "all the data a user needs is not on display at each step in a transaction sequence,"
7. "prompts, cues, and/or messages are not placed where the eye is likely to be looking on the screen,"
8. "there is missing information or explanations," and
9. "meaningful groups of items are not separated by white space" (see Appendix A: Heuristic Evaluation).

In addition, the online library catalog had twenty-two items that were rated as major usability issues. These items included:

- “every page display does not begin with a title or header that describes the screen contents,”
- “continuity of thinking is required by needing to remember information through several screens,”
- “high levels of concentration necessary for remembering information from screen to screen,”
- “it is not relatively easy for the user to understand where they might wish to go next in the system” and
- “the system does not provide navigational aids for the users as they navigate between multiple screens.”

The context-rich hypermedia learning environment had twenty items that were rated as major usability issues. These items included:

- “after the user completes an action (or group of actions), feedback does not indicate that the next action can be started,”
- “the system does not provide navigational aids for the users as they navigate between multiple screens,”
- “it is not relatively easy for the user to understand where they are in the system,”
- “buttons are not adequately labeled” and
- “each window does not have a title.”

It is important to note that all catastrophic and major usability issues were mitigated for both hypermedia environments except in the case of the online library catalog when to do so would have meant that the display and/or functionality would no longer have mirrored that of the Library of Congress' Online Catalog (LOC Catalog) or the Michigan State University's MAGIC online catalog (MAGIC) (see "Conditions" below). For example, the catastrophic item identified for the online library catalog, "images, text and audio files/sounds are not easy to access" concerned the fact that resources did not automatically launch or begin playing when the evaluator was at the record level but required one additional click. However, this design is consistent with both the LOC Catalog and MAGIC. An additional example of a major usability item identified for the online library catalog was "labeling and language are not clear for each record," which concerned the fact that the terms "heading" (from "subject heading") and "type" were confusing to some of the evaluators. In the case of "subject heading" the label was shortened to "subject," which was consistent with MAGIC. However, as both the LOC Catalog and MAGIC use the term "type," this label was retained. In the case of the context-rich condition button, labels were enhanced and each window was given a title.

Conditions

The purpose of the study was to measure how differences in the conditions affected participants' recall and knowledge integration. The control condition (Figure 3-2. Linear/ control condition) used a minimal and linear hypermedia structure consisting of several design features that were antithetical to those of the

experimental condition and replicated the functionality¹⁵ of the Library of Congress' Online Catalog and Michigan State University's MAGIC online catalog. Specifically, in the control condition individuals had to search for the information needed, locate a relevant bibliographic record (

Figure 3-3. Linear/ control bibliographic record) and then click within the object to access the actual information. Each bibliographic record provided the type of information (e.g., text, audio and images), title, author, description, date and subject as well as a direct link to the information itself.¹⁶ The control condition was intended to present the *same resources* as the experimental condition, but in a rigid and decontextualized manner. It was postulated that the linear interface would foster greater mastery of factual knowledge (e.g., an object's title).

¹⁵ Current as of March 2006.

¹⁶ The Library of Congress' Online Catalog and Michigan State University's MAGIC online catalog contain direct links to the information itself only when the resource is available electronically or when some portion such as a table of contents or publisher's description is available. The vast majority of content described in traditional library catalogs is not available electronically.

Online Library Catalog: Flint Sit-Down Strike

[Home](#)
[Hours](#)
[Requests](#)
[Help](#)

The Online Library Catalog is a database of cataloging records representing the collection of Flint resources held by the Library. These resources are composed of text, audio and images, which can be searched by 'Type of Material'.

Search By:

Search Examples

Title	Title Example: "Working Conditions Caused the Strike"
Author	Author Example: "Joe Fry" or "Fry, Joe"
Keyword	Keyword Example: "violence," "wages," "women"
Type of Material	Type of Material Example: "text," "image," or "audio"
Subject Heading	Subject Example: "Working Conditions - Michigan - Flint"

Figure 3-2. Linear/ control condition

Online Library Catalog: Flint Sit-Down Strike

New Search

Hours

Requests

Help

Type: audio

Title: Store Keepers and Farmers Attitudes Toward the Strike

Author:

- ♦ Fry, Joe

Description: Fry talks about store keepers who were against the union and wouldn't give any food. He also tells about a sympathetic farmer who brought in a whole hog to the strike kitchen.

Date: 1979-07-27

Subject:

- ♦ Community Response - Michigan -- Flint
- ♦ General Motors Corporation Sit-Down Strike, 1936-1937
- ♦ Conditions During Strike - Michigan -- Flint

Connect to: Store Keepers and Farmers Attitudes Toward the Strike

Title

author

keyword


Type of Material

Subject Heading

Figure 3-3. Linear/ control bibliographic record

The experimental condition (

Figure 3-4. Context-rich/ experimental condition) incorporated design features derived from cognitive flexibility theory, including context-dependency and interconnectedness, which acknowledge and manage complexity (Spiro et al., 1992). In particular, several presentations of the same information in multiple contexts were provided to highlight different facets of the information. In addition, the experimental condition displayed an interwoven series of images, text and audio to form a context-rich representation. Specifically, individuals browsed thematic areas (though they could also search as well) to locate a related informational object and could immediately see or hear the information (Figure 3-5. Complex rich/ experimental resource) while discovering the type of information (e.g., text, audio, and images), title, author, description, date and “theme” for the information. This condition was intended to present the same resources as in the control condition but in a cognitively flexible format that was hypothesized to enhance the ability of learners to synthesize and integrate new knowledge.



Digital Library: Flint Sit-Down Strike

[Home](#) |
 [Working Conditions Before the Strike](#) |
 [Strike Methodology](#) |
 [Conditions During the Strike](#) |
 [Community Response to the Strike](#)

keyword

[FAQ](#) | [Sealpage Guide](#)

Working Conditions Before the Strike:

Provides an overview of the employment conditions for thousands of General Motors workers that lead to the decision to sitdown or occupy the plants rather than simply leave and use a conventional walkout strike.

Strike Methodology:

Provides an overview of the confusion and decision-making surrounding the start of the strike for both the workers inside the plants as well as people who weren't.

Conditions During the Strike:

Provides an overview of both the day-to-day activities strikers and non-strikers were involved in such as getting food in and keeping the police out.

Community Response to the Strike:

Provides an overview of how the people of Flint and Michigan, supporters and non-supporters, viewed the strike. Reactions were mixed, but opinions were strong.



Working on the line at General Motors in Flint was a job many men needed desperately in the 1930's, but it was also tremendously difficult. Terrible working conditions, combined with unfair and devious payroll practices, made the auto plants of Depression-era Flint Michigan into ripe locations for union organization.

[Home](#) |
 [Working Conditions Before the Strike](#) |
 [Strike Methodology](#) |
 [Conditions During the Strike](#) |
 [Community Response to the Strike](#)

Figure 3-4. Context-rich/ experimental condition

keyword

[FAQ](#) | [Realplayer Guide](#)

This record resulted from a search for leo

Wages and Breaks at General Motors

11

Type: audio

Title: Leo Connelly was threatened by the general foreman

Author:

- Connelly, Leo

Description: His brother convinced him to come to Flint to work in the Buick Plant. Connelly started at fifty cents an hour and never got a raise. After four years, he quit to go to Chevrolet plant where the pay was higher. He discusses the injustices of the bonus system. When he had his wages cut and wasn't able to get an answer about it, he angered the general foreman by going to the main office. Connelly says they treated workers "just like a dog" before the strike.

Date: 1980-03-06

Subject:

- Working Conditions &c Michigan -- Flint
- General Motors Corporation Sit-Down Strike, 1936-1937

[Home](#) | [About this site](#) | [Contact Us](#) | [Privacy Policy](#) | [Terms of Use](#) | [Feedback](#)

Figure 3-5. Complex rich/ experimental resource

It was postulated, based on the findings of Jacobson and Spiro (1995) that the context-rich interface, which supported simultaneity and interconnections, would increase overall study time but also facilitate greater understanding and knowledge integration. Resources for both of the systems were composed of audio files, images and texts. Audio files comprised slightly fewer than 60% of the resources, images just over 30% of the resources and texts just over 10% of the resources (Table 3-5. A breakdown of the audio files, images and texts in both conditions).

Table 3-5. A breakdown of the audio files, images and texts in both conditions by theme area

	Audio	Images	Texts	Total
Working Conditions	16	5	2	23
Strike Methodology	8	5	2	15
Conditions During the Strike	9	7	2	18
Community Response	10	5	2	17
TOTAL audio, images & texts	43	22	8	73

The control and experimental condition contained the same 73 resources, and each of these resources was categorized in exactly the same manner. For example, any resource that had the subject heading “working conditions” in the control condition was located in the “working conditions” theme in the experimental condition. Nine resources (audio, text and images) that were relevant or interconnected to more than one theme appeared in two themed areas and one resource (an image) that was relevant or interconnected to more than one theme appeared in three themed areas (Table 3-7. A numeric breakdown of unique and shared resources by theme). For example, the audio resource entitled “Food sources during the strike,” which described who in the community assisted the strikers by providing food, was relevant to “conditions during the strike” and “community response.” The control and experimental condition each contained 56 unique resources. Of these unique resources, 33 were comprised of audio, 17 of images and 6 of texts (Table 3-6. A numeric breakdown of total and unique resources).

Table 3-6. A numeric breakdown of total and unique resources

Type of resource	Number of resources
Total audio resources	43
<i>Total unique audio resources</i>	33
Total images	22
<i>Total unique images</i>	17
Total texts	8
<i>Total unique texts</i>	6

Table 3-7. A numeric breakdown of unique and shared resources by theme area

	One	Two	Total
	Theme	Themes	
Working Conditions	20	3	23
Strike Methodology	9	6	15
Conditions During the Strike	12	6	18
Community Response	13	4	17
TOTAL unique & shared resources	54	19 ¹⁷	72

¹⁷ Though this column sums to 19, a total of 10 resources are available in two or more different themed areas.

The audio files average 72.72 seconds (i.e., just over one minute) per file with a range of 212 seconds (i.e., three and a half minutes) to 25 seconds (i.e., less than one half minute) and with a standard deviation of 44.1 (see Figure 3-6. Distribution of the length of sound files).

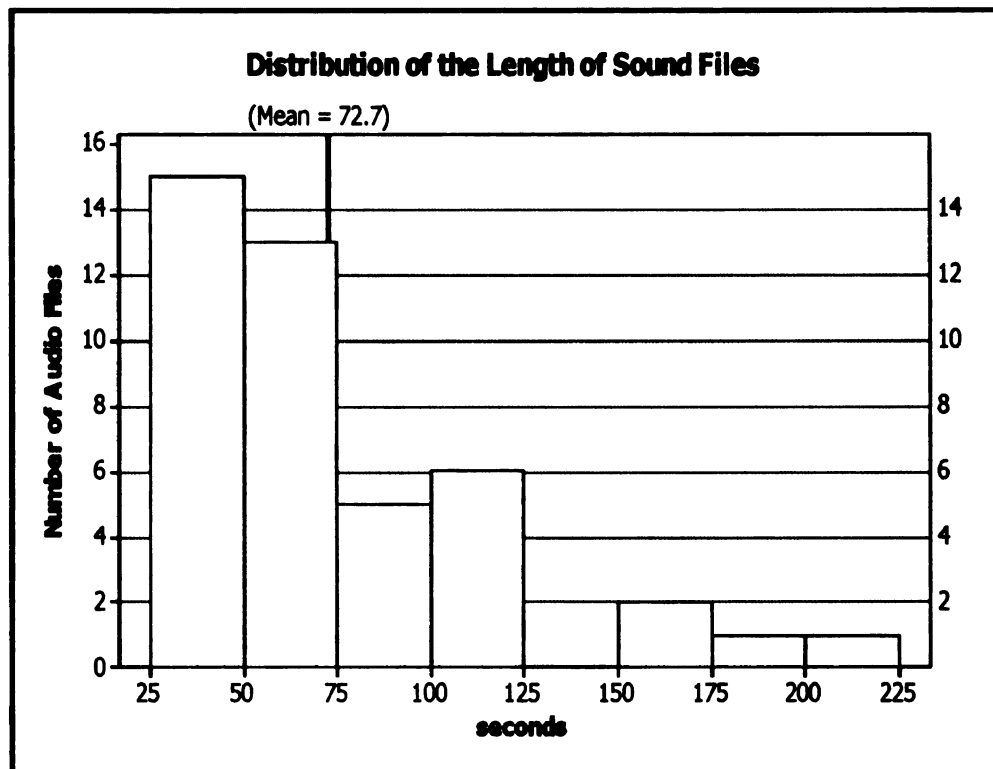


Figure 3-6. Distribution of the length of sound files

The text files average 874 words per file with a range of 653 words to 1122 words and with a standard deviation of 172.31 (see Table 3-8. Number of words in Text Files).

Table 3-8. Number of words in Text Files

Themed area	First text file	Second text file
Working Conditions	870 words	1122 words
Strike Methodology	653 words	1122 words
Conditions During the Strike	921 words	765 words
Community Response	774 words	765 words

The only content change aside from the different interfaces (Figure 3-2. Linear/ control condition as contrasted with

Figure 3-4. Context-rich/ experimental condition) was an introductory paragraph for each theme in the experimental hypermedia learning environment. These paragraphs ranged from 111 words to 136 words (see Table 3-9. Word count for introductory paragraphs).

Table 3-9. Word count for introductory paragraphs

Theme	Number of words
Working Conditions	136 words
Conditions During the Strike	111 words
Strike Methodology	133 words
Community Response	116 words

Pilot Testing

Prior to data collection pilot data were collected in an effort to determine how procedures and data collection would work in practice. A convenience sample of five participants was recruited via a College of Education graduate student mailing list. All participants were taking doctoral courses or had completed their dissertations but had not yet graduated. Three participants were female and two were male. Three participants had lived all or most of their lives in the U.S. while two had come to the U.S. to go to graduate school. Three participants were assigned to the control condition and two were assigned to the experimental condition. Each participant completed part one, the initial testing and learning stage, and part two, the factual recall, synthesis and integration test stage except for the vocabulary test. The retention stage was added after pilot testing was completed (See Figure 3-7. Pilot Test Testing Flowchart).

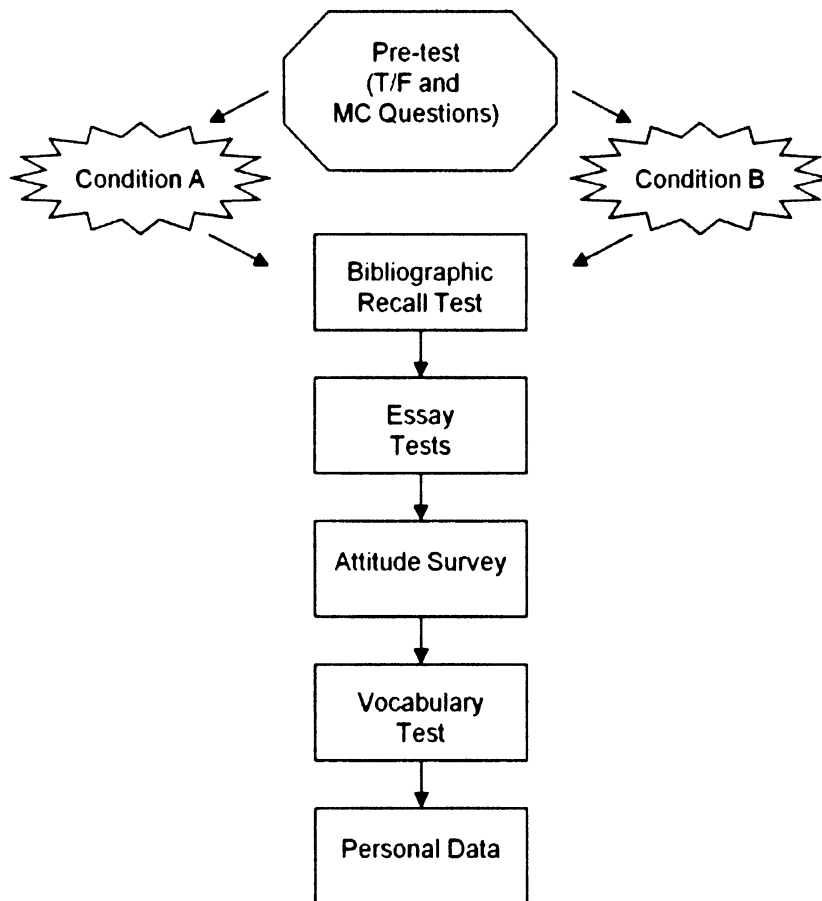


Figure 3-7. Pilot Test Testing Flowchart

Although thorough statistical analyses were not appropriate, time spent in each stage, select test results and attitudes were analyzed. The time participants spent in part one, which included the background test for Content Knowledge and System Exploration, ranged from 35 to 96 minutes ($\mu = 62.6$). The time participants spent in part two, which included the Evaluation Measures, Attitude Survey and Personal Data Collection, ranged from 13 to 60 minutes ($\mu = 33.8$). Though there was a great deal of variance in the time each participant spent on

parts one and two, the mean times were used to estimate the approximate amount of time the actual test participants would need.

Pre-tests for content knowledge ranged from 4 to 6 correct out of a possible score of 8 correct. Though the participants were all highly educated people these results were better than expected given that two of the five participants were not from the U.S. Therefore, a 9th question, which was judged more difficult, was added for the actual test participants. The goal of the additional question was to allow participants with a greater knowledge of Michigan labor history to demonstrate *more* content knowledge. The Evaluation measures (Recall and Essay tests) were not scored. They were used, however, to develop a rubric for grading the actual test participant essays. (A complete list of measures and data collected follows.)

The Attitude survey showed some differences between the two conditions. Pilot participants in the linear, control condition assigned higher scores to the following statements: "I liked the Flint Sit-Down Strike system," "I would recommend this system to other students," and "I learned a lot about the Flint Sit-Down Strike using this system." Whereas pilot participants in the context-rich, experimental condition assigned higher scores to the following statements: "The system was easy," "It was easy to understand where I was in the system," "It was easy to find information in the system" and "It was easy to move around in the system". The Personal data collected revealed that all had used a library catalog such as MAGIC before to find books magazines or journals, checked out at least 4 books during the recently completed semester and spent at least 6 hours per

week searching for information (rather than reading email) on the Internet including Google.com, Amazon.com, Magic (MSU Library catalog), etc.

Lab and Equipment

Participants were tested in the Berkey Hall, Room 216 (see Appendix O: Berkey 216) and the Kedzie Hall South, Room 222 (see Appendix N: South Kedzie 222) computer labs at Michigan State University. As Berkey Hall is where the majority of history classes are taught, it was expected that the computing lab would be relatively easy for participants to locate and the site itself might be perceived as more comfortable because of its familiarity. However, the Berkey Hall lab has very restricted hours. Kedzie Hall, a nearby lab, had less restricted hours so it was also used. Both PC labs were equipped with 32 Dell Pentium 4 or faster processors, 256 MB RAM, at least 80 GB computers operating Windows XP Pro and using 15" multi-scan monitors. The computers were arranged in rows of 2 to 5. Headphones were provided, which allowed the individual participants to discretely hear the audio while not disturbing anyone around them.

Testing Procedures

Procedural Overview

Testing was divided into three stages. In stage one, the initial testing and learning stage participants completed the pre-test for content knowledge and used their assignment condition. In stage two, the factual recall, synthesis and integration test stage, participants responded to a series of test questions with short answers and essays. In stage three, and one week after the intervention,

participants' achievement was again measured via the completion of a test for retention. (See Figure 3-1: Testing flowchart).

Upon arriving at the lab, the test participants were asked to read, ask questions about and sign informed consent forms (Appendix H: Informed Consent and Explanation Form). Following this initial step, they drew at random a unique number (between 1 and 36). Participants were identified by this unique number throughout testing in order for responses from stages one and two to be linked anonymously with responses from stage three. Next, they were read and provided with a copy of the general instructions (Appendix I: General Instructions). Finally, they accessed the survey instrument via SurveyMonkey (<http://www.surveymonkey.com/s.asp?u=717383072837>) where they logged all responses under their unique number. After completing the pre-test for background knowledge (Appendix A: Content knowledge Flint Sit-Down Strike) they received specific instructions on paper (Appendix J: Specific Instructions) to the system which they were randomly assigned (as a group¹⁸), as well as a list of all the resources available (Appendix K: Resources). After they finished using the system, they completed the evaluation measures (Appendix C: Evaluation Measures), and vocabulary test (Ekstrom et al, 1976) and answered a few brief personal questions (Appendix F: Personal Data). One week following the intervention, participants' achievement was again measured via the completion of a test for retention (Appendix M: Post Test - Content knowledge of the Flint Sit-Down Strike). (See **Error! Reference source not found.**). In general, testing

¹⁸ Testing sessions were organized with between 6 to 2 participants. See "Sample" for more.

time for stage one, the initial testing and learning stage, was completed in less than 110 minutes.

Table 3-10. Time Required	
System Exploration	≤ 60 minutes
Appendix C: Evaluation Measures	≤ 30 minutes
Appendix D: Attitude Survey	≤ 5 minutes
Vocabulary Test	≤ 8 minutes
Appendix F: Personal Data	≤ 2 minutes

Following the study participants were debriefed as to the exact nature of the study (Appendix L: Debriefing Form), asked if they had any questions and/or if they would like additional information.

Content Knowledge

Participants were pre-tested for knowledge of the Flint Sit-Down Strike via eight true/false or multiple choice questions (Appendx B: Content Knowledge Flint Sit-Down Strike). Participants were asked to identify via true/false questions whether “the decision to strike was made by the majority of unionized employees” and whether “the National Guard was deployed to protect both parties from harming one another.” Participants were asked to identify, using multiple choice responses, the relative time frame of the strike, the automobile manufacturer involved, the factors that contributed to the strike, the novel approach used by the strikers, the manufacturer’s stance toward unionization

and, more generally, national labor leaders. These questions focused on basic facts and concepts. They were designed to enable participants with lesser knowledge of Michigan labor history to demonstrate *some* content knowledge and those with greater knowledge of Michigan labor history to demonstrate *more* content knowledge within a 5-minute period. The question concerning national labor leaders helped to identify those participants who have some labor history knowledge but weren't necessarily knowledgeable about Michigan labor history. This helped to "remove" prior knowledge from dependent criterion test results. These questions were developed by the author and Rand Spiro, Professor, Counseling, Educational Psychology, & Special Education, and reviewed by Dr. Dale Belman, Professor, Labor and Industrial Relations at Michigan State University.

Evaluation Measures

Following the pre-test, participants received a list of instructions (Appendix I: General Instructions and Appendix J: Specific Instructions) and a list of all the resources that were available (Appendix K: Resources) and then explored the condition to which they were randomly assigned. They were asked to listen to exactly 20 audio recordings and read exactly 4 text files and to use their Resources handout to either check off each resource and/or to keep a running total. This was done to try to ensure that any differences in achievement were not a function of the number of information resources to which participants were exposed. There was no minimum or maximum amount of time they were required to spend in their condition. This was done to try to ensure that participants had

sufficient time to explore their condition as well as listen to and read the appropriate information resources. Though participants were permitted to take pen and paper notes during the System Exploration phase on the list of instructions (Appendix I: General Instructions and Appendix J: Specific Instructions) and/or the list of all the resources (Appendix K: Resources) that were available, all handouts were collected prior to testing.

After system exploration, participants' were tested on dependent Evaluation Measures (Appendix C: Evaluation Measures). Next, their attitude toward the system to which they were assigned was surveyed (Appendix D: Attitude Survey). Participants were timed on how many minutes they spent exploring their system and how many minutes they spent on testing.

Recall, Synthesis and Integration

Participants were tested for recall, synthesis, integration and retention. Participants' memory of the objects themselves was measured by requesting that they provide title, author, and/or subject heading for any six audio files, texts and/or images. In addition, participants were asked to write two essays containing a minimum of 250 words each. The first tested knowledge integration across themes. This essay question was, "Keeping in mind working conditions, strike methodology, conditions during the strike, and community responses, how did General Motors try to control its workers?" The second tested knowledge synthesis within a specific theme. The question, developed with Michael van

Dyke¹⁹, was, "Using specific examples, what were people's rationales in their stance for or against the strike?" Lastly, knowledge retention was studied by asking participants about key concept to which they had been exposed one week following exposure. Inter-rater reliability for the assignment of scores was assessed (See "Scoring Procedures" for more).

Recall was assessed in order to measure how much bibliographic information (type of resource, title, author, description, date and subject headings) could be remembered. In addition, it was assessed in order to determine how many discrete resources participants remembered. Recall is an important measure in hypermedia research on structural differences as a means of examining format effects on learning (e.g., Chen, Ghinea, & Macredie, 2006; Brunye', Taylor, Rapp & Spiro, 2006; Lee & Tedder, 2003; Brinkerhoff, Klein & Koroghlanian, 2001; Dee-Lucas & Larkin, 1999; Jonassen, 1993). It was also ecologically valid as Marchionini notes because use of an online catalog necessarily relies on recall knowledge. Users must think of words from memory to enter into the catalog (as cited in Borgman, Hirsh, Walter & Gallagher, 1995). Further, individuals can encounter quite a bit of information in the process of retrieving results. Being able to recall pertinent specifics aids the research process and is a measure of engagement since participants must focus their attention on particular elements (Bransford, Brown, & Cocking, 2000).

The first essay question, which tested knowledge integration, could only be answered effectively after examining resources from the system as a whole

¹⁹ Michael van Dyke was a Visiting Assistant Professor, Department of American Thought and Language who worked extensively with the audio files and authored several short essays on the Strike.

(i.e., “working conditions,” “strike methodology,” “conditions during the strike” and “community response”). In answering this question directly after exposure to their assigned condition, without further intervention (e.g., test questions), participants needed to integrate different information from the system as a whole in order to prepare the most effective response.

The second essay question, which tested knowledge synthesis, could be answered after examining the resources labeled “Working Conditions” in the control condition or the resources grouped in “Working Conditions before the Strike” in the experimental condition. This meant that participants did not need to integrate information from either system as a whole. As Figure 3-2. Linear/ control condition shows, the example “Working Conditions” is provided on the control condition home page. It was also provided in both sets of instructions to participants (Appendix I: General Instructions). In addition, as Table 3-5. A breakdown of the audio files, images and texts in both conditions shows, out of the four categorized or themed topics, the slight majority of resources are categorized as “Working Conditions” (i.e., 23 resources). This meant that participants in either condition had a slightly better than 1 in 4 chance of being exposed to resources related to “Working Conditions” and therefore, differences in performance could be attributed to condition, not content.

Attitude survey

Participants’ attitudes toward their respective conditions were surveyed via 7 statements using a 6-point Likert scale ranging from 1 (strongly agree) to 6 (strongly disagree) (Appendix C: Attitude Survey). Statements focused on ease

of use and general satisfaction with their respective condition. Examples of statements include "I liked this system," "I would recommend this system to other students," "I learned a lot about the Flint Sit-Down Strike using this system," and "It was easy to find information in the system." The survey also included two constructed response items that asked participants what they liked best and least about the system.

Though attitude was not one of the key dependent criteria, it was hypothesized that differences based on assignment to condition would be found. This is significant because just as it is important to know if learning is occurring and what types of learning are occurring in different learning environments, it is also important to know any barriers to the use of different learning environments.

Vocabulary Testing

Participants were tested for verbal comprehension via a two-part advanced Vocabulary Test that is part of the Kit of Factor-Referenced Cognitive Tests (Ekstrom et al, 1976). Specifically participants had a total of 8 minutes to select, via multiple choice, synonyms for 36 multi-syllabic words. For each vocabulary word, one of five responses had to be chosen (e.g., Mumble: speak indistinctly, complain, handle awkwardly, fall over something, tear apart). The test was designed for students in grades 11-16. The Manual for Kit of Factor-Referenced Cognitive Tests (Ekstrom, French & Harman, 1976) states that "...research has suggested that verbal comprehension is a sub factor involving reading comprehension, verbal analogies, matching proverbs, grammar and syntax" (163). It was therefore important to understand and control for the effect of an individual

participant's verbal comprehension in order to separate cognitive ability and effect of the condition.

Personal Data

Data on gender and personal characteristics were also collected. Participants were asked to report their gender. Participants were asked whether they had ever used a library catalog such as MAGIC before to find books, magazines or journals. Participants were asked how many books, journals, etc. they had checked out from a library in the most recently completed semester. The scale was: 0-3, 3-6, 6-10, 10-15 and 15 or more. Participants were asked how many hours per week they spent searching for information (rather than reading email) on the Internet, including on search sites such as Google.com, Amazon.com, and Magic (MSU Library catalog). The scale was 0-3 hours, 3-6 hours, 6-10 hours, 10-15 hours, and 15 or more hours (see Appendix F: Personal Data).

Retention Measure

Participants were asked to complete a test for retention (see Appendix M: Post Test - Content knowledge of the Flint Sit-Down Strike) one week after exposure to either condition. Participants were given up to 48 hours to complete this test from any location they wished. They logged responses using the number (between 1 and 36) that they had previously drawn at random in order to successfully compare their responses to previous test data. The retention test required fewer than 10 minutes to complete and was composed of 10 questions

and one short answer question designed to determine participants' factual recall and synthesis of the Flint Sit-Down Strike content. Two of the questions were identical while, two more were very similar (see Table 3-11. Similar questions on the Pre and Post Tests), one was narrower and two were new questions.

Table 3-12. Similar questions on the Pre and Post Tests	
Pretest	Post Test
<p>The major factors that contributed to the strike were:</p> <ul style="list-style-type: none"> • Reduced retirement packages • Wages and working conditions • Reduced health care benefits 	<p>The most common complaint that striking Flint workers voiced was that GM:</p> <ul style="list-style-type: none"> • Refused to provide medical insurance • Did not provide a pension plan • Pushed constantly to speed-up production
<p>The company's stance toward unionization was:</p> <ul style="list-style-type: none"> • One of partnership • One of caution • One of opposition 	<p>Prior to the Flint sit-down strike, General Motors had:</p> <ul style="list-style-type: none"> • Welcomed the UAW in its plants • Attempted to halt union membership drives in its plants • Engaged with the UAW in collective bargaining

The short-answer question asked participants to “Imagine that you are an employee in one of GM’s Flint plants. A fellow worker approaches you and asks you to participate in the sit-down strike that has just broken out in Fisher One and Two. Would you join in the strike? Using specific examples, name three reasons why you would or why you would not.” These questions were designed to test for basic and more advanced facts and concepts.

Scoring Procedures

Participants did not disclose their identity during testing but did use a randomly assigned number (between 1 and 36) to log data in stages 1 through 3. All data was gathered via SurveyMonkey.

The Content Knowledge (Appendix A) assessed knowledge of the Flint Sit-Down Strike via eight true/false or multiple-choice questions. Responses were scored as correct or incorrect. Questions one through seven were each worth one point while question eight, which asked “which five of the following people gained fame as national labor leaders,” was worth a total of five or one point per person correctly identified.

The Evaluation Measures included tests for recall, synthesis and integration (Appendix C: Evaluation Measures). The Recall test (Appendix C: Evaluation Measures, Part I) asked participants to write down the title, author and/or subject of six of the resources (audio, text and/or image). The Recall test was scored as correct or incorrect. If any of the information provided by a participant for a title, author and/or subject could be used to retrieve one or more resources then the test question was scored as correct. If resources could not be

retrieved from the information supplied by the participant then the test question was scored as incorrect. The highest possible score was a 6 and the lowest possible score was a 0. For example, while “eating candy on the line” would retrieve a resource, “eating on the line” would not.

The essay test for Integration (Appendix C: Evaluation Measures, Part II) was evaluated and subsequently scored based on four discrete components: the number of themes discussed or mentioned, the number of facts discussed or mentioned, the extent to which themes and facts were connected coherently in the writing and the overall score. The highest possible score was a 12 and the lowest possible score was a 0 (see Figure 3-8. Integration Measure Rubric).

Figure 3-8. Integration Measure Rubric	
Criteria	Score
I. Themes: refers to the number of themes explicitly mentioned in the essay. (4 to 0 points)	4: 4 themes mentioned 3: 3 themes mentioned 2: 2 themes mentioned 1: 1 themed mentioned 0: 0 themes mentioned
II. Facts: refers to the number of facts explicitly mentioned in the essay ²⁰ . (4 to 0 points)	4: ≥ 10 facts mentioned 3: 7 – 9 facts mentioned 2: 4 – 6 facts mentioned 1: 1 - 3 facts mentioned 0: 0 facts mentioned

²⁰ Any factual errors need to be deducted at a 1:1 ratio from the score given for Facts. For example, if a participant mentions 7 facts with one misstatement, then the total number of facts 6 leading to a total score of 2. An example of a misstatement is “(the) heat killed many workers.”

Table 3.8 (cont'd).

<p>III. <i>Connections</i>: refers to the extent to which themes and facts are connected coherently in the writing.</p> <p>(4 to 0 points)</p>	<p>4: Two or more themes and four or more facts for each theme are clearly connected within a coherently structured essay. No more than one misrepresentation is present in the essay.</p> <p>3: Themes and facts are connected within a somewhat coherently structured essay (e.g., person may discuss working conditions, community response and working conditions again).</p> <p>2: One or two themes and/or fewer than four facts for each theme are loosely connected within a poorly structured essay.</p> <p>1: Themes and facts are not developed and not well connected within a poorly structured essay (e.g., themes are introduced without supporting facts).</p> <p>0: Themes and facts are not presented.</p>
---	--

The Integrative essay question was: "Keeping in mind working conditions, strike methodology, conditions during the strike, and community responses, how did General Motors try to control its workers?" Supportive examples that participants might have given included: by attempting to move work out of Flint just prior to the strike, by using violence (i.e., the Flint police) against the strikers during

previous strike attempts, by turning off heat to the building occupied by the strikers (in January), by stationing the National Guard outside the building occupied by the strikers, as well as those supportive examples cited for the second essay, which follow.

The essay test for Synthesis (Appendix C: Evaluation Measures, Part III) was evaluated and subsequently scored based on one component: the number of facts discussed or mentioned. The highest possible score was a 6 and the lowest possible score was a 0 (see Table 3-13. Synthesis Measure Rubric.)

Table 3-13. Synthesis Measure Rubric	
Criteria	Score
<i>I. Facts:</i> refers to the number of facts explicitly mentioned in the essay ²¹ . (6 to 0 points)	6: ≥ 16 facts mentioned 5: 13 – 15 facts mentioned 4: 10 – 12 facts mentioned 3: 7 – 9 facts mentioned 2: 4 – 6 facts mentioned 1: 1 -3 facts mentioned 0: 0 facts mentioned

The Synthesis essay question read, “Using specific examples, what were people’s rationales in their stance for or against the strike?” Supportive examples that might have been included were: poor working conditions, poverty wages (e.g., “In 1935, a year in which the government declared \$1,600 as the minimum income

²¹ Any factual errors need to be deducted at a 1:1 ratio from the score given for Facts. For example, if a participant mentions 7 facts with one misstatement, then the total number of facts is 6 leading to a score of 4. An example of a misstatement is “(the) heat killed many workers”.

on which a family of four could live decently, the average auto worker took home \$900”), unequal pay, job insecurity, and layoffs. Specific examples of poor working conditions include not being able to take breaks to use the bathroom, get water and eat; high indoor temperatures (e.g., Maynard Mundale passed out while working the line); the difficulty in keeping up with the line work; having to work for bosses on their cabins or sidewalks and most of all the work speed-up.

Alternatively, people might have noted that occupying the facility was trespassing, a company union was sufficient to represent workers’ interests, a strike might undermine the local economy or that striking was undemocratic.

The Attitude Survey (Appendix D: Attitude Survey) responses ranged from Strongly agree (6), Agree (5), Weakly agree (4), Weakly disagree (3), Disagree (2), Strongly disagree (1). The Vocabulary Test (Ekstrom et al, 1976) responses were scored as correct or incorrect and provided inferential statistics. Personal information (Appendix F: Personal Data) was used to determine general population characteristics, use of any physical libraries, and time spent in any information seeking activities.

The essay test for Retention (Appendix M: Post Test - Content knowledge of the Flint Sit-Down Strike) was evaluated and subsequently scored based on one component: the number of facts discussed or mentioned. The highest possible score was a 4 and the lowest possible score was a 0 (see Table 3-14. Retention Measure Rubric.)

Table 3-14. Retention Measure Rubric	
Criteria	Score
<i>I. Facts:</i> refers to the number of facts explicitly mentioned in the essay. ²² (4 to 0 points)	4: ≥ 4 facts mentioned 3: 3 facts mentioned 2: 2 facts mentioned 1: 1 fact mentioned 0: 0 facts mentioned

Supportive examples that might be included in the Retention measure are the same as those that might be included in the Synthesis measure.

Amy Tracy Wells, the examiner, independently scored the Content Knowledge (Appendix A) and Vocabulary (Appendix D) tests, which could be scored as correct or incorrect. The examiner and Dale Belman, Professor, Labor and Industrial Relations at Michigan State University, scored the same sub-set of randomly chosen essay responses. Specifically, a total of eight essay responses for the integration measure, a total of eight essay responses for the synthesis measure and eight of the short-answer responses for the retention measure were scored (i.e., 24 responses that comprised slightly more than 25% of all essay and short-answer responses) by two scorers. Differences in coding were discussed and either maintained or resolved. For those essays measuring integration, there was an inter-coder agreement of 91.66%. For those essays measuring synthesis, there was an inter-coder agreement of 100%. For those short-answer responses

²² Any factual errors need to be deducted at a 1:1 ratio from the score given for facts. For example, if a participant mentions 7 facts with one misstatement, then the total number of facts is 6 leading to a score of 4. An example of a misstatement is "(the) heat killed many workers."

measuring retention, there was an inter-coder agreement of 87.5%. Following this high rate of inter-rater reliability (i.e. a Pearson correlation of $>.80$), the examiner scored the remainder (Howell, 2002).

Statistical Tests

Difference in means tests and multiple regression using ordinary least squares was used to analyze the data and determine if conditions are associated with differences in achievement and other outcomes. These tests were used to determine if there are systematic differences in outcomes between the two conditions. As participation in the two groups was randomized, t-tests were theoretically sufficient to determine if there were differences in outcomes between the two conditions on the population. However, as the samples were small, randomization may not have sufficiently controlled for the effects of achievement, verbal comprehension or time spent using a condition. Regression models were used to control for factors other than condition that might systematically affect outcomes. Descriptive statistics were also provided.

CHAPTER 4. STATISTICAL ANALYSIS

Participants' responses to both interface conditions were measured using difference in mean tests and regression. Difference in means tests measured whether the difference in performance between the groups was statistically significant and whether such differences could be expected between the populations studied. Similarly, regression analysis tests for statistically significant differences in the performance between the participants in the experimental and control conditions but controls for factors that might mask the experimental effect. For example, regression allows control for factors such as gender, content knowledge, vocabulary comprehension etc., thereby removing any masking effects of these factors from the participants' performance. Therefore, regression tests are important for understanding the significance of individual differences on performance.

Throughout this chapter, results for explanatory variables are reported using one- or two-tailed tests as appropriate. For example, in regression tests that analyze the performance of individuals with higher scores on the background and vocabulary tests and in which one might assume, all else were constant, that these individuals would perform better on outcome measures, a one-tailed test was used. However, when there were no such presumptions, for example in tests that analyze performance by gender, two-tailed test were used.

The key finding from both the difference in means tests and regression tests is that there was little difference in the performance of the two groups. There was some evidence that the control condition performed better on the

factual recall measure, made more connections and retained for facts on the test for retention but the *overall* difference in performance of the two groups was never statistically significant. This outcome was not the result of imprecision related to the use of relatively small samples ($n = 16$ and $n = 17$); rather, the magnitude of the difference in performance between the two groupings was consistently small. This suggests that the two systems are similar in their support of learning.

Financial and temporal considerations limited the number of respondents who were tested, and as a result, the experimental and control samples are small. Small samples are characterized both by larger variances than larger samples from the same population and, if the samples are sufficiently small, a non-normal distribution of the sample test statistics. Because of these two factors, it is more difficult to reject the null hypothesis of no effect than would be the case with a larger sample. A .05 test may, therefore, be too strict a standard for not rejecting the null for this size sample. In order to better protect against failing to reject the null in error, a .05 and .10 standard is applied. The failure of even a .10 test to reject the null in our tests provides stronger evidence that the difference in performance between the experimental and control group is not meaningful. The section entitled "Implications of Sample Size for these Findings" discusses this issue further.

Survey I

Data were collected from 33 participants who were randomly assigned to the experimental or control condition (see Appendix P: Specifics of Each

Participant). Seventeen participants were assigned to the control condition and 16 to the experimental condition. Of the participants in the control condition, 8 were male and 9 were female. Of the participants in the experimental condition, 8 were male and 8 were female. In all, 9 test sessions were held with 1 to 7 participants ($\mu = 3.66$) per session.

Background Performance Measures

The survey included several batteries of questions to determine participants' knowledge of labor history, attitudes about the condition to which they were assigned vocabulary, and use of libraries and familiarity with searching for information. They provide evidence that the experimental and control groups were similar with respect to specific knowledge and cognitive ability thereby supporting the use of t-tests for differences in means. These same variables are used as controls for the regression analysis.

Content Knowledge

Participants' knowledge of labor history was evaluated with eight True/False and Multiple Choice questions. Seven were general questions about labor history. The eighth asked participants to identify five labor leaders from a list of ten historic and fictional individuals. As the latter question was particularly challenging, question eight was analyzed separately from the first seven questions.

Flint Sit-Down Strike questions:

The assessment of participants' knowledge of labor history included factual questions such as, "The Flint Sit-Down Strike took place in {Michigan in the 1930s, 1950s or 1970s}" and "The company's stance toward unionization was {One of partnership, One of caution or One of opposition}". Correct answers received 1 point, while incorrect answers received 0 points. The responses to these questions were summed. The scores potentially ranged from 0 (no correct answers) to 7 (all correct); but the actual range of scores for the 33 participants was 1 to 6. The 33 participants averaged 3.80 correct answers with a standard deviation of 1.4 and a median of 4 correct answers.

The response did not vary systematically by condition. Those in the control group averaged 3.6 correct responses, while those in the experimental group averaged 4.1 correct responses. The medians were 3 and 4 respectively. A t-test for a difference in means between the two groups could not reject a two tailed null of no difference in a .05 test ($t[29] = 0.95, p = .175$).

Labor leader recognition:

Participants had considerable trouble differentiating historic labor leaders from other historic figures and fictional figures. Participants averaged 1.9 correct responses with a standard deviation of .89 correct answers and a median two correct answers. Differences in the number of correct responses between participants in the control and experimental group were modest. The control averaged 2.1 correct responses, the experimental group averaged 1.6, and the median for both groups was 2. A t-test for the difference in means between the

two groups could not reject the null of no difference in a two-tailed .05 percent test ($t[31] = -1.645, p = .117$).

Combined Measure:

The two measures of background knowledge of labor history were aggregated into a single measure assigning a value of 1 to each correct answer. This meant that questions 1-7 were each worth one point and question 8 was worth 5 points. The highest possible score was a 12 and the lowest possible score was a 0.

The mean for the combined measure was 4.2 with standard deviation of .25 and median of 4.2. The mean for the control group was 4.0, while that for the experimental group was 4.4; the medians were 3.6 and 4.4 respectively. A t-test for a difference in means cannot reject the null of no difference in .05 percent test ($t[31] = 0.76, p = .773$).

The two types of questions were designed to measure different levels of knowledge. That is, it was designed to enable participants with lesser knowledge of Michigan labor history to demonstrate *some* content knowledge and those with greater knowledge of Michigan labor history to demonstrate *more* content knowledge. The simple correlation of the labor history and labor leader recognition measure is $-.14$, but it is far from statistically significant ($t[32] = -.14, p = .435$) suggesting that they measure different aspects of individuals knowledge of labor history.

Vocabulary

Intellectual ability was measured by a vocabulary test that required the respondent to determine the best synonym from five possible choices for each of thirty-six words. The Vocabulary test is part of the Kit of Factor-Referenced Cognitive Tests. Since "...verbal comprehension is a sub factor involving reading comprehension, verbal analogies, matching proverbs, grammar and syntax" (Ekstrom, French & Harman, 1976, 163), it was therefore important to control for the effect of an individual participants' verbal comprehension to distinguish cognitive ability and effect of the condition. The 33 participants averaged 22.5 correct responses; the standard deviation was 5.3 responses. The median respondent had 22 correct responses; the distribution of responses is right skewed.

There were no statistically significant differences between the control and experimental groups. The control had 20.2 correct answers on average, while the mean for the experimental group was 21.5. The medians were 19.0 and 22.0 respectively. A t-test for a difference in the means of the control and experimental group failed to reject the null in a .05 or .10 test ($t[31] = 0.67, p = .25$).

Personal information

Correlation of Knowledge of Labor History Measures and Vocabulary Measures:

The correlation between the vocabulary and knowledge of labor history measures is small ($R_2 = .278, p = .117$). Therefore, it is not possible to reject the null of no correlation in either a five or ten percent test. The small magnitude of

the correlation and the inability to reject the null suggests that these two measures are capturing different aspects of individuals' background and abilities.

Attitude

Toward the end of the survey, participants were asked to evaluate the system they were using with respect to ease of use, their understanding of materials, ease of finding information, movement around the system and whether they would recommend the system. Responses were scored on a six point Likert scale running from "strongly like" to "strongly dislike".

Most participants reported high levels of satisfaction with both systems. The aggregate scores for the seven items ranged from a low of 4.76 to a high of 5.2 without regard to whether they were assigned to the experimental or control condition. The median response for each item was 5.0, suggesting that the majority participants gave the high rating to the system, but a few participants gave relatively low ratings. This issue is further discussed after a comparison of the response by condition.

Responses to the Attitude Questions by Condition:

Data on the means, standard deviations, medians, minimum and maximum by condition, and a hypothesis test for differences in response by experimental condition, for the attitude variables are provided in Figure 4-1. Attitude toward the System (1) and Figure 4-2. Attitude toward the System (2). The means and medians for each condition are quite similar for each item and it is not possible to reject the null of no difference by condition in a .05 t-test for any of seven items. For example, the mean for the experimental condition for whether

the individual would recommend the system is 4.97, the mean for the control condition is 4.95. The median for both conditions is 5.0. A hypothesis test for a difference in the response by experimental condition has a p-value of .54, ($t[31] = 0.11$, $p = .54$) well above the level required to reject the null in any conventional hypothesis test. Similar results were obtained for each of the other six attitude variables. The difference in these values range from .1 to .3, but hypothesis tests for a difference did not come close to rejecting the null of no difference in a .05 test. For example, the strongest t-value was 0.58 with a corresponding p-value of .57. Parallel results were obtained when the values of the seven attitude variables are summed. The mean value for full sample was 4.92, the means for the control and experimental condition were 4.93 and 4.95 respectively. Again, a t-test for a difference in means is not able to reject the null of no difference ($t[31] = 0.11$, $p = .54$). In sum, there is little evidence of difference in participants' perception of the two systems.

Responses within Group:

The histograms shift our focus from comparing the means between the two conditions, to considering the distribution of responses within condition. For each item and within each condition, the majority of respondents provide answers of 5 or 6 for each question. However, there is distinct minority who is dissatisfied or who had trouble with the system. For example, there was a 2, a 3 and two 4's in response to the question would you recommend this system in the experimental condition; and a 2 and two 4's in the control condition (see Figure 4-1. Attitude toward the System (1)). As can be seen in response to the

questions "I learned a lot" there were also differences within condition (see

Figure 4-1. Attitude toward the System (1)).

There was very high correlation between some items. For example, the three questions measuring ease of use (i.e., "It was easy to understand where I was in the system," "It was easy to find information in the system," and "It was easy to move around in the system") had correlations ranging from .402 to .657 and all correlations were significant in better than a 2 percent test (see Figure 4-2. Attitude toward the System (2)). Likewise, there was a strong correlation between the two questions measuring whether or not the participants would recommend the system and how much they felt they learned (i.e., "I would recommend this system to other students" and "I learned a lot about the Flint Sit-Down Strike using this system") had a correlation of .462 that was significant in better than a 1 percent test ($p=0.007$) (see Figure 4-1. Attitude toward the System (1)).

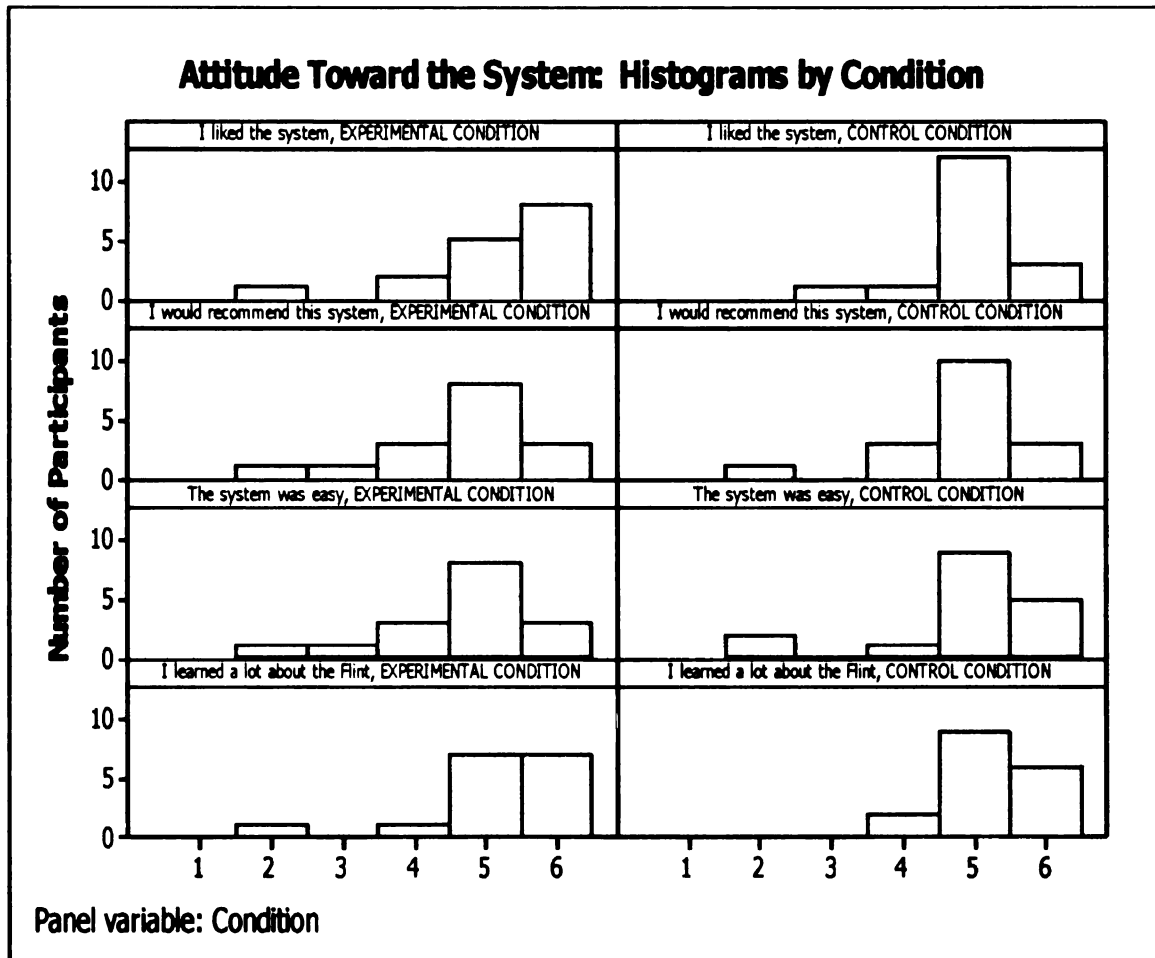


Figure 4-1. Attitude toward the System (1)

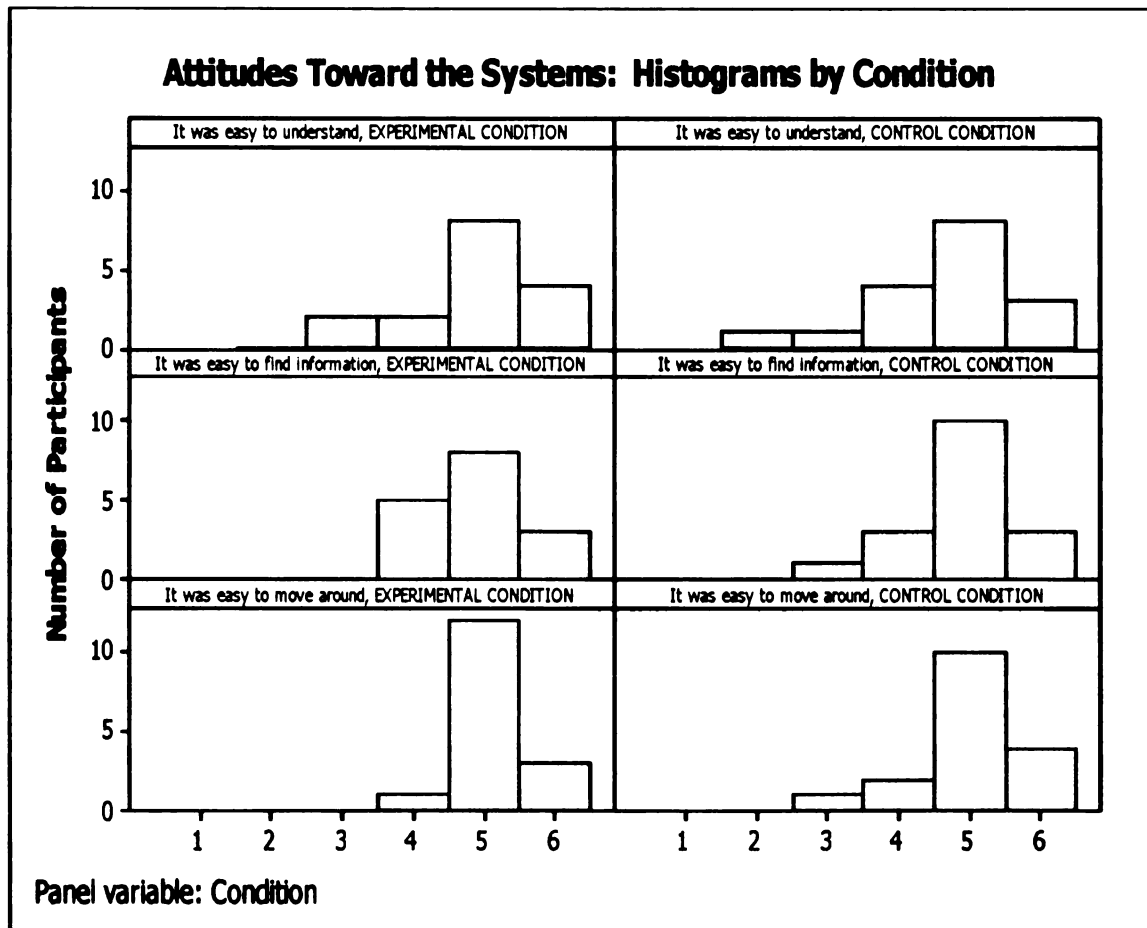


Figure 4-2. Attitude toward the System (2)

Regression Analysis of Attitude Questions

A regression analysis of the attitude questions was conducted focusing on the total attitude measure. The regression included controls for gender, content knowledge and vocabulary comprehension, library catalog use, the amount of materials checked out from a library in the most recently completed semester and the amount of time spent searching the Internet (see Table 4-1. Regression on Attitudes) as well as an 0/1 indicator variable for whether the respondent was in the experimental or control condition. Inclusion of these controls removes any

masking effects these factors may have on the relationship between condition and attitudes. The inclusion of gender as a control reflects the ubiquity of the influence of gender found in a broad range of studies of cognition and learning. The remaining controls, which measure experience with library systems and the internet, control for differences in attitudes related to prior general experience with the types of systems and resources used in this study. The total attitude variable has a mean of 4.935, a standard deviation of .096 and ranges from 3.857 to 6.0.^{23 24}

. The R^2 for this model was 10.1 percent. Controlling for these additional factors does not affect the measured relationship between participants' condition and

²³ Use of regression to remove masking effects may be complicated by the presence of multicollinearity. Correlation between variables increases the estimates sensitivity to sample differences and increases the variance of estimated coefficients. A test for multicollinearity using a variance inflation factor was conducted and each of the seven explanatory variables was regressed on the remaining six explanatory variables. The values of the variance inflation factors were all close to 1, ranging from 1.065 to 1.209, so multicollinearity is unlikely to be an issue in this data set. As the same explanatory variables are used throughout the study, there are no further discussions of multicollinearity, however R^2 statistics are provided.

²⁴ The number of degrees of freedom in the regression equation varies from 21 to 25, depending on the equation under consideration. Although the small number of degrees of freedom raises issues of the accuracy of the hypothesis test, it is possible to guard against the tendency to not reject the null in small samples by using a weaker standard for rejection, .10 rather than .05 test. Despite this weakening of the standard for rejection of the null of no effect, there is never any evidence that the experimental condition out performs the control condition.

their overall attitude ($\beta=0.06$, $t=0.28$, $p=0.782$). None of the added variable are statistically significant in a .05 or .10 two tailed test in their own right.

Table 4-1. Regression on Attitudes

Predictor	Coef	SE Coef	T	P (two tailed)
Constant	5.050	0.686	7.37	0.000
Experimental Condition	0.061	0.218	0.28	0.782
Total Background Score	0.086	0.080	1.08	0.291
Total Vocabulary Correct	-0.010	0.023	-0.45	0.657
Male	-0.220	0.211	-1.04	0.309
Library Catalog Use	0.083	0.647	0.13	0.899
No. books checked out	-0.044	0.127	-0.35	0.728
Hours spent searching	-0.069	0.101	-0.69	0.497
S = 0.592472 R-Sq = 10.1% R-Sq(adj) = 0.0%				

* significant in a 10 percent test or better

Participants were also asked what “the best system of the program was” (see Table 4-2. Rank Ordering of Responses to “The best part of the (Control) system was...”). This question was open-ended. In the control condition, the most common responses were (in descending order)

1. “ease of use including navigation”,

2. that the system was composed of “multimedia” and “audio” and
3. the system’s “organization”.

In the experimental condition, the most common responses were (in descending order) that the system

1. included “audio”, “primary...” and “multimedia” resources,
2. its organization, and
3. included text files.

It is interesting to note that while access to multimedia and audio resources were identified as “the best part of the system” by participants in both conditions, twice as many participants in the experimental condition did so than in the control condition. It is also interesting to note that participants in the control condition most often identified “ease of use including navigation” as “the best part of the system”.

Table 4-2. Rank Ordering of Responses to “The best part of the (Control) system was...”

Response	Number of respondents
Ease of use including navigation	6
Multimedia	3
Audio	3
Organization	3
Primary resources	1
Aided the development of “analytical connections”	1
Novel	1
Different opinions	1
Aided the development of a “solid understanding about a historical event”	1
Resemblance to a “library system”	1
Simple and non-distracting pages	1

Table 4-3. Rank Ordering of Responses to “The best part of the (Experimental) system was...”

Response	Number of respondents
Audio	8
Primary resources	5
Multimedia	4
Organization	3
Text files	2
Resource descriptions	1

Participants were also asked what “the most difficult part of the system was” (see Table 4-4. Rank Ordering of Responses to “The most difficult part of the (control) system was...”). This question was open-ended. In the control condition, the most common responses were (in descending order):

1. “difficulty searching”,
2. “poor audio quality” and
3. that the “resource types difficult to determine”.

Two participants responded however that there were no problems. In the control condition, the most common responses were (in descending order)

1. “poor audio quality”
2. “duplicated/repetitive resources”,
3. “poor organization audio or text” and
4. that “some audio was too brief or the content not helpful”.

It is interesting to note that more participants in the experimental condition found “poor audio quality” and “duplicative/repetitive resources” problematic than participants in the control condition.

Table 4-4. Rank Ordering of Responses to “The most difficult part of the (control) system was...”

Response	Number of respondents
Difficulty searching	7
Poor audio quality	2
Resource types difficult to determine	2
No problem	2
No background information on the Strike itself provided	1
Duplicated/repetitive resources	1
Too many results when searching	1
Amount of clicking (in order to access a resource)	1
Insufficient information about the resources	1
Organization	1
Insufficient information on the resources	1

Table 4-5. Rank Ordering of Responses to “The most difficult part of the
(experiment) system was...”

Response	Number of respondents
Poor audio quality	6
Duplicated/repetitive resources	4
Poor organization audio or text	2
Some audio was too brief or the content not helpful	2
Insufficient variety in the resources (e.g., more radio broadcasts and newspaper accounts)	1
Insufficient information on the authors and speakers	1
“Reading some of the articles and some of the audio files”	1
“Understanding some of the audio monologues”	1
No problems	1
Browser crashed	1
“Retaining specific items in individual memory such as title, author, etc... Subject matter was strongly retained though”	1

It should be noted that while two participants stated that they had trouble determining resource types, all resources did in fact indicate whether they were

text, audio or image so the issue may have had more to do with whether participants read this information. In addition, while the comments “reading some of the articles and audio responses” and “understanding some of the audio monologues” might be interpreted as “poor audio quality”, it wasn’t clear if this was the case so each was listed separately.

Recall

Participants recall of discrete aspects of the resources and their labeling was evaluated with six fill-in-the-blank questions. Specifically participants were asked to “Write down as much as you recall about one of the Flint resources (audio, text and/or image) such as a title, author and/or subject”. This question sought to measure how much bibliographic information could be recalled and to determine how many discrete resources participants recalled as measured by how much bibliographic information they were able to remember correctly.

Responses were scored as correct or incorrect based on whether or not their response could be used to locate the resource if typed into a keyword (unquoted) search²⁵. Participants received full credit if they correctly named the title, author or subject of a resource even though they might not be able to remember the full title, author or subject of the resource (e.g., “Van Dyke” for “Michael Van Dyke” or “Burning burlap” for “Burning burlap instead of coal”). In addition, participants received full credit if the additional words added did not affect retrieval (e.g., “Wages and Wage Conditions in the GM Plants” for “Wages

²⁵ Typos and minor punctuation errors were ignored (e.g., “Police protection for the Judges family” for “Police protection for the Judge’s family”)

and Breaks at General Motors"). However, responses such as "Radical leadership" instead of "Strike Leaders Were More Radical" wherein a suffix has been added (i.e., "ship") and the words re-ordered was scored as incorrect since the item could not be located using a keyword search and the meaning of the title was different from the resource itself.

As a whole:

The thirty-three participants were asked to supply up to three pieces of information for up to six resources. Correct information was provided for 57 titles, 29 authors and 33 subjects, meaning that out of a possible 594 possible responses, participants were able to correctly recall 118 titles, authors and/or subjects. Participants had one of three general responses when asked to "write down as much as you recall about...subject". Either they provided correct subject information (e.g., "Strike methodology"), a close approximation (e.g., "Wages and working conditions" for "Working Conditions"), or they described a text, audio or image resource (e.g., "A company stockholder visits the worker occupied plants. He sympathizes with the workers after he discovers that they've been protecting company property and that the police caused most of the damage".) Subjects or close approximations (e.g., "Working conditions before the strike" for "Working conditions") were provided for 42 resources while descriptions were provided for 95 resources. While the question sought to determine whether subjects could be recalled, the descriptions are important and are briefly discussed below.

Some responses included (correct) title or author information *and* (correct) subject information but not for the same resource. The responses were scored as

correct if the subject was plausible (e.g., “Mrs. Moon” and “Conditions during the Strike” for “Community Response”). One response included (correct) author *and* a description (not a subject) but not for the same resource. This response was scored as incorrect as the directions requested information about “...one of the Flint resources...” and a singular resource could not be distinguished.

Consistent with the hypothesis, the control group ($\mu = 4.53$ vs. $\mu = 2.56$) scored considerably higher on recall than did the experimental group ($t[31] = 2.01$, $p = 0.027$). This difference in means translates in the control group having an average recall of an additional 1.97 items. The null of no difference in recall between the experimental and control group was rejected in both a .10 and .05 one tailed test.

This result is also supported by the regression analysis. Following the prior specification, this analysis included controls for scores on content knowledge, vocabulary comprehension, library catalog use, the amount of materials checked out from a library in the most recently completed semester and the amount of time spent searching the Internet as well as gender. The R^2 for the model is 41.9 percent. Factors significantly affecting recall include vocabulary and gender. Participants who scored higher on the vocabulary items had higher recall scores ($\beta=0.291$, $t=3.01$, $p=0.003$); men performed worse on the recall items than did women ($\beta=-1.55$, $t=-1.70$, $p=0.101$). The null of no difference could be rejected in a .05 and 0.10 test respectively. Similar to the difference in means test, the regression indicates that participants in the experimental group had lower recall scores ($\beta=-2.00$, $t=-2.14$, $p=0.021$) than did

those in the control group. The coefficient ($\beta=-2.00$) on experimental condition indicated that the experimental participants gave two fewer recall items than did those in the control group. The null of no difference could be rejected in both a .10 and .05 one tailed test.

Table 4-6. Regression on Recall				
Predictor	Coef	SE Coef	T	P (two tailed)
Constant	0.491	2.952	0.17	0.869
Experimental Condition	-2.009	0.940	-2.14	0.042*
Total Background score	0.211	0.345	0.61	0.547
Total Vocabulary correct	0.291	0.097	3.01	0.006*
Male	-1.550	0.910	-1.70	0.101*
Library Catalog Use	-0.505	2.786	0.18	0.858
Number books checked out	0.309	0.545	-0.57	0.576
Hours Info Search	0.387	0.433	-0.89	0.380
S = 2.55043 R-Sq = 41.9% R-Sq(adj) = 25.7%				
* significant in a 10 percent test or better				

Descriptions of Essays

Participants were also tested on their ability to synthesize and integrate information via two essay questions. The first explored knowledge integration across themes via an essay containing a minimum of 250 words that each

participant was asked to write. This essay question was, "Keeping in mind working conditions, strike methodology, conditions during the strike, and community responses, how did General Motors try to control its workers?" The second explored knowledge synthesis via an essay containing a minimum of 250 words that each participant was asked to write. The question was, "Using specific examples, what were people's rationales in their stance for or against the strike?"

The first essay question, which tested knowledge integration, could only be answered effectively after examining resources from the system as a whole (i.e., "working conditions," "strike methodology," "conditions during the strike" and "community response"). By asking this question directly after exposure to their assigned condition, participants without further intervention (e.g., test questions) needed to integrate different information from the system as a whole in either condition in order to prepare the most effective response. These essays were scored separately on the number of themes in the essay, the number of facts in the essay, the connections between themes and facts, and an overall score (which was calculated as the sum of the three dimensions) on which the essay was scored. Two participants were removed from this analysis because their essays were less than seventy words. However, one participant's essay contained 240 words but was retained. The average word count for the essays as a whole was 274.12. Examples of these essays are located in Appendix E: Exemplar Integration Essays.

The second essay question, which tested knowledge synthesis, could be answered after examining the resources labeled "Working Conditions" in the

control condition or the resources grouped in “Working Conditions before the Strike” in the experimental condition. This meant that participants did not need to integrate information from either system as a whole. These essays were scored based on the number of facts included in the essay. Two participants were removed from this analysis because their essays were less than seventy words. However, two participants’ essays contained 208 and 239 words but were retained as the number of facts included in each, 4 and 3 respectively, were close to the average score of 3.74 with a mode 4. The average word count for the essays as a whole was 293.45. Examples of these essays are located in Appendix E: Exemplar Synthetic Essays.

Integration

There are four outcome measures for the integrative essay: the number of themes, the number of facts, the connections between the themes and facts, and the overall score. The scoring on themes is a simple count of the number of themes identified in the essay. The measures of facts and connections are ordered responses corresponding to the level of performance of the respondent. For example, the measure of facts was scored on a scale of 0 to 4 with 0 corresponding to no facts, 1 corresponding to 1 – 3 facts, 2 corresponding to 4 -6 facts, 3 corresponding to 7 – 9 facts and 4 corresponding to 10 or more facts mentioned. The measure of connections was also scored from 0 to 4 with 0 corresponding to an essay with no discernable themes and facts, 1 corresponding to an essay in which themes and facts are not developed and not well connected within a poorly structured essay, 2 corresponding to an essay in

which one or two themes and/or less than four facts for each theme are loosely connected within a poorly structured essay, 3 corresponding to an essay in which themes and facts are connected within a somewhat coherently structured essay and 4 corresponding to an essay in which two or more themes and four or more facts for each theme are clearly connected within a coherently structured essay containing no more than one misrepresentation. The overall score was the sum of the scores on themes, facts and connections. The highest possible score was a 12 and the lowest possible score was a 0 (see Table 4-7. Overall means). The typical integrative essay included a discussion of two and three themes (see Table 4-8. Integration Rubric), included between seven and nine facts (and subsequently received a score of 3), included between two and three connections and received an overall score of between eight and nine.

Table 4-7. Overall means

Measure	Highest possible score	Mean	Std error
Number of Themes	4	2.613	0.110
Number of Facts	4	3.419	0.159
Connections	4	2.774	0.172
Total Integration Score	12	8.806	0.298

Table 4-8. Integration Rubric

Score	Explanation
4	Two or more themes and four or more facts for each theme are clearly connected within a coherently structured essay. No more than one misrepresentation is present in the essay.
3	Themes and facts are connected within a somewhat coherently structured essay (e.g., person may discuss working conditions, community response and working conditions again).
2	One or two themes and/or less than four facts for each theme are loosely connected within a poorly structured essay.
1	Themes and facts are not developed and not well connected within a poorly structured essay (e.g., themes are introduced without supporting facts).
0	Themes and facts are not presented.

Analysis of the sample by condition suggests that a participant's condition did not affect their performance (see Table 4-9. Integration Essay).

Table 4-9. Integration Essay

Measure	Condition	Number of respondents	Mean	Std. Dev.
Number of themes				
	Control	17	2.588	0.173
	Experimental	14	2.643	0.133
Number Facts				
	Control	17	3.294	0.239
	Experimental	14	3.571	0.202
Connections				
	Control	17	2.882	0.241
	Experimental	14	2.643	0.248
Total Integration Score				
	Control	17	8.765	0.407
	Experimental	14	8.857	0.455

The difference in items group means are consistently minimal. The difference in the mean number of themes, .055, did not achieve statistical significance in a .one tailed .05 or .10 test ($t[29] = 0.25$, $p = .40$). Similarly, the difference in the number of facts on the integrative essay was .28 but non-significant in a .05 or .10 one tailed test ($t[29] = 0.89$, $p = .19$); the difference in connections was -.239 and both wrong signed and non-significant ($t[29] = -0.69$,

$p = .75$); and the difference in overall score was similarly small, .092, and non-significant in a .05 or .10 one tailed test ($t[29] = 0.15$, $p = .56$).

The measures of number of themes, number of facts, number of connections and overall score were then regressed on controls for the measures of content knowledge, vocabulary comprehension, gender, library catalog use, the number of books checked out from a library in the most recently completed semester and the amount of time spent searching the Internet. R^2 greater than 40 percent for three of the four models. However, the indicator variable for experimental condition was not correctly signed in any model and not significant in a 5 or 10 percent one tailed test. In the total integration regression, the coefficient for experimental condition was negative and the t-test has a probability of .44; the coefficient in the number of facts regression was also negative and had a two tailed p-value of .997; the coefficient in the number of connections regression was also negative, rather than the hypothesized positive, and did not achieve significance in a .05 or .10 test ($\beta = -0.4634$, $t = -1.52$, $p = 0.141$).

Considering the other explanatory variables in the regressions, scores on the vocabulary test had a strong positive relation to the integrative connections score and the total integrative score ($\beta = 0.174$, $t = 2.79$, $p = 0.010$). Gender, specifically being male ($\beta = 0.659$, $t = 2.48$, $p = 0.021$), and having used a library catalog ($\beta = 1.92$, $t = 2.47$, $p = 0.021$) had a positive relation to the number of facts cited.

Table 4-10. Regression on Total Integration Score

Predictor	Coef	SE Coef	T	P (two tailed)
Constant	3.349	1.683	1.99	Blank
Experimental Condition	-0.422	0.540	-0.78	0.443
Total Background Score	-0.152	0.204	-0.74	0.465
Total Vocabulary Correct	0.174	0.062	2.79	0.010*
Male	0.470	0.536	0.88	0.390
Library Catalog Use	1.820	1.568	1.16	0.258
Number of books checked out	0.311	0.324	0.96	0.347
Hours spent searching	-0.053	0.286	-0.19	0.853
R-Sq = 42.8% R-Sq(adj) = 25.4%				

* significant in a 10 percent test or better

Table 4-11 Regression on Number of Themes

Predictor	Coef	SE Coef	T	P (two tailed)
Constant	2.832	0.772	3.67	0.001
Experimental Condition	0.043	0.248	0.17	0.864
Total Background Score	-0.078	0.094	-0.83	0.416
Total Vocabulary Correct	0.012	0.029	0.44	0.667
Male	-0.123	0.247	-0.50	0.622
Library Catalog Use	-0.530	0.720	-0.74	0.469
Number of books checked out	0.184	0.149	1.23	0.230
Hours spent searching	-0.131	0.141	-0.10	0.921
R-Sq = 12.2% R-Sq(adj) = 0.0%				

* significant in a 10 percent test or better

Table 4-12. Regression on Number of Facts

Predictor	Coef	SE Coef	T	P (two tailed)
Constant	0.391	0.833	0.47	0.643
Experimental Condition	-0.001	0.267	-0.00	0.997
Total Background Score	-0.051	0.101	-0.51	0.618
Total Vocabulary Correct	0.042	0.031	1.36	0.186
Male	0.659	0.266	2.48	0.021*
Library Catalog Use	1.921	0.777	2.47	0.021*
Number of books checked out	-0.114	0.161	-0.71	0.486
Hours spent searching	0.182	0.141	1.29	0.210
R-Sq = 50.7% R-Sq(adj) = 35.7%				

* significant in a 10 percent test or better

Table 4-13. Regression on Connections

Predictor	Coef	SE Coef	T	P (two tailed)
Constant	0.126	0.948	0.13	0.896
Experimental Condition	-0.463	0.304	-1.52	0.141
Total Background Score	-0.023	0.115	-0.20	0.844
Total Vocabulary Correct	0.120	0.0352	3.40	0.002*
Male	-0.067	0.302	-0.22	0.828
Library Catalog Use	0.429	0.884	0.49	0.632
Number of books checked out	0.242	0.183	1.32	0.199
Hours spent searching	-0.222	0.161	-1.38	0.180
R-Sq = 45.2% R-Sq(adj) = 28.5%				

* significant in a 10 percent test or better

Synthesis

There was one outcome measure for the synthesis essay: the number of facts supplied. The highest possible score was a 6 and the lowest possible score was a 0 (see Table 4-14. Synthesis scores by condition.) The typical synthesized essay received a score of between three and four²⁶. The mean response for the full sample was 3.74. The mean for the control condition was 3.64 and the mean for the experimental condition 3.85. A test for a difference in means cannot reject

²⁶ A score of three or four means that the participants included 7-9 facts or 10 – 12 facts respectively.

the null of no difference in a .05 or .10 test of significance ($t[29] = 0.42$, $p = .34$). As with other tests, the lack of significance is not due to large standard errors, but rather to the outcomes for the two groups being notably similar.

Table 4-14. Synthesis scores by condition

	Number	Mean	P
Total Synthesis Score	31	3.742	0.245
Control condition	17	3.647	0.342
Experimental condition	14	3.857	0.361

Control for other factors does not materially alter the conclusion that the experimental condition did not affect the total recall score. The coefficient is small in magnitude ($\beta=0.007$, $t=0.01$, $p = 0.989$) and is far from achieving significance in a one tailed .05 or .10 test. R^2 for the model is 17.7 percent.

Table 4-15. Regression on Total Synthesis

Predictor	Coef	SE Coef	T	P (two tailed)
Constant	2.568	1.659	1.55	0.135
Experimental Condition	0.007	0.532	0.01	0.989
Total Background Score	-0.204	0.201	-1.02	0.320
Total Vocabulary Correct	0.0305	0.062	0.50	0.624
Male	-0.515	0.529	-0.97	0.340
Library Catalog Use	0.533	1.546	0.34	0.734
Number of books checked out	0.281	0.320	0.88	0.389
Hours spent searching	0.194	0.281	0.69	0.497
R-Sq = 17.7% R-Sq(adj) = 0.0%				

* significant in a 10 percent test or better

Survey II: Post test

Data was collected from 31 of the 33 initial participants. Responses to the Post-test True/False and Multiple Choice questions ranged from 5 to 7 correct ($\mu = 5.71$) out of 8 questions. The Post-test question of "which five of the following people were involved in the Flint Sit-Down Strike" was difficult for participants. Responses ranged from 1 to 4 correct ($\mu = 2.65$). A similar Pre-test question, "which five of the following people gained fame as national labor leaders," was also difficult for participants and responses ranged from 0 to 3 correct ($\mu = 1.9$).

Content Knowledge

There is little difference in the means for either the T/F or Labor leaders questions in either condition. The mean number correct for T/F questions was 5.58 for the experimental condition and 5.82 for the control condition. The mean number of labor leaders correct was 2.36 for the experimental condition and 2.9 for the control condition. In neither instance was the difference statistically significant in a one tailed .05 or .10 test ($t[29] = -0.50$, $p = 0.689$ and $t[29] = -1.62$, $p = 0.94$) respectively.

Table 4-16. Content knowledge

	Number	Mean	Std error	S.D.
T/F Total Correct	31	5.710	0.246	1.371
Total Correct Labor Lead	31	2.645	0.164	0.915

Table 4-17. True/False Correct

	Number	Mean	Std error	S.D.
Control	17	5.824	0.324	1.334
Experimental	14	5.571	0.388	1.453

Table 4-18. Labor Leaders Correct

	Number	Mean	Std error	S.D.
Control	17	2.882	0.208	0.857
Experimental	14	2.357	0.248	0.929

Somewhat different results were obtained from the regression analysis. The Post test's T/F and Labor Leader scores were regressed on the same set of explanatory variables used in the prior regression equations. Although the adjusted R^2 for the equations is 8.7 and 0.0 percent respectively for the T/F and Labor Leader questions, the experimental condition indicator variable is correctly signed and significant in a one tailed .05 test in the labor leader equation ($\beta = -0.678$, $t = -1.87$, $p = 0.074$). The experimental condition variable is not significant in the T/F equation ($\beta = -0.348$, $t = -0.68$, $p = 0.25$). The result for the labor leader equation is consistent with the hypothesis about the effect of hypermedia systems on fact retention. Specifically, that participants in the linear, control condition would achieve higher scores than participants in the context-rich, experimental condition on the test for factual recall.

In addition to the results for the experimental condition variable, vocabulary scores have a positive effect in both equations – achieving a .10 but not a .05 level of significance in both equations. The background score has a positive effect on the T/F response ($\beta = 0.315$, $t = 1.71$, $p = 0.051$), but is not significant for the labor leader equation ($\beta = -0.059$, $t = -0.45$, $p = 0.327$).

Table 4-19. Regression of T/F Post Test

Predictor	Coef	SE Coef	T	P (two tailed)
Constant	4.256	1.537	2.77	0.011
Experimental Condition	-0.349	0.512	-0.68	0.502
Total Background Score	0.315	0.185	1.71	0.101*
Total Vocabulary Correct	0.071	0.052	1.36	0.186
Male	0.679	0.484	1.40	0.174
Library Catalog Use	-1.059	1.440	-0.74	0.469
Number of books checked out	-0.235	0.288	-0.82	0.422
Hours spent searching	0.031	0.228	0.13	0.895
R-Sq = 30.0% R-Sq(adj) = 8.7%				

* significant in a 10 percent test or better

Table 4-20. Regression of Labor Leaders Post test

Predictor	Coef	SE Coef	T	P (two tailed)
Constant	1.412	1.090	1.30	0.208
Experimental Condition	-0.679	0.363	-1.87	0.074*
Total Background Score	-0.059	0.131	-0.45	0.654
Total Vocabulary Correct	0.052	0.037	1.42	0.168
Male	0.013	0.343	0.04	0.970
Library Catalog Use	0.905	1.021	0.89	0.384
Number of books checked out	-0.150	0.204	-0.74	0.468
Hours spent searching	0.057	0.162	0.35	0.727
R-Sq = 21.0% R-Sq(adj) = 0.0%				

* significant in a 10 percent test or better

Retention

Participants were tested on their ability to retain the information they were exposed to via a short response. This measure did not specify the number of words required but did ask participants to identify three facts. The question was

Imagine it is the winter of 1936-37 and you are an employee in one of GM's Flint plants. A fellow worker approaches you and asks you to participate in the sit-down strike that has just broken out in Fisher One and Two. Would you join in the strike? Using specific

examples, name three reasons why you would or why
you would not?

The highest possible score was a 4 and the lowest possible score was a 0 (see Table 4-21. Retention Scores.) Examples of scored essays are located in Appendix E: Exemplar Retention Essays.

Thirty-one of the initial grouping of thirty-three participants participated. Two participants though they responded to the T/F and multiple choice questions did not provide responses to the short answer question meaning that a total of 29 short answers were evaluated or approximately 89% of the original participants responded.

The average score on this essay was 2.9, the mean for the experimental group was 2.79 and the mean for the control group was 3.07. A t-test for the difference in these means does not reject the hypothesis that the population means are no different in a .05 or .10 test ($t[27]=-0.56$, $p=0.708$)

Table 4-21. Retention Scores

	Number	Mean	Std error	S.D.
Retention Score	29	2.931	0.248	1.334
Retention Score Control	15	3.067	0.316	1.223
Retention Score Experimental	14	2.786	0.395	1.477

Contrary to the difference in means tests, regression of the retention score on experimental status and the measures of content knowledge, vocabulary comprehension, gender, familiarity with library catalogs, the amount of materials

checked out from a library in the most recently completed semester and the amount of time spent searching the Internet indicates that being in the control condition is weakly associated with better performance on retention ($\beta=-0.698$, $t=-1.36$, $p=0.094$). The R^2 for the equation is 33.4 percent, adjusted R^2 is 11.2 percent. The score on the background questions has a positive and strongly statistically significant ($\beta=0.411$, $t=2.29$, $p=0.017$) effect on retention.

Table 4-22. Regression of Total Retention

Predictor	Coef	SE Coef	T	P (two tailed)
Constant	-0.192	1.519	-0.13	0.900
Experimental Condition	-0.698	0.514	-1.36	0.189*
Total Background Score	0.411	0.180	2.29	0.033*
Total Vocabulary Correct	-0.008	0.054	-0.14	0.889
Male	-0.509	0.489	-1.04	0.310
Library Catalog Use	1.647	1.388	1.19	0.248
Number of books checked out	0.103	0.285	0.36	0.722
Hours spent searching	0.143	0.221	0.65	0.526
R-Sq = 33.4% R-Sq(adj) = 11.2%				

* significant in a 10 percent test or better.

Implications of Sample Size for these Findings

One issue with this analysis is whether the lack of statistical significance is due to small sample sizes, or because of the small absolute difference in the group means. Using the sample means and standard deviations for the population values, and assuming that the ratio of sample group sizes would be the same in the population, finding a statistically significant result for the total integrative measure that would reject the null in a .05 one tailed test and have a power of 0.9 would require an control sample of 422 respondents and 348²⁷ experimental respondents. Modest increases in the size of the sample would then likely not affect the measured outcomes. Rather, the means of the experimental and control groups are, in a statistical sense, similar.

Did Learning Occur?

A threshold issue is whether learning occurred with either of the systems. This can be tested with a one sample t-test against a null that the difference in pre-test and post test score was zero. The mean difference in Content Knowledge scores from Pre-test to Post-test for the 31 respondents to the post-test was 1.97 for the T/F and 0.74 for labor leaders. Both differences were statistically significant in both a .10 and .05 one tailed test (for T/F: $t[29]=6.84$, $p=0.00$; Labor Leaders: $t[29]=3.77$, $p=0.00$).

Similar results were obtained for learning for each of the two conditions. With respect to the T/F questions, the mean for the experimental condition was 1.64 ($t[13]=4.25$, $p=0.00$) while the mean for the control condition was 2.24 ($t[16]=5.37$, $p=0.00$) (see Table 4-23. Descriptive Statistics Pre-test to Post-test

27 These statistics were calculated and contributed by Professor Dale Belman.

and Table 4-24. T/F Experimental and Control Condition). Therefore, the null of no learning could be rejected in a .10, .05 and .01 test. With respect to the Labor Leader questions, the means for the experimental and control conditions were 0.71 ($t[13]=2.22$, $p=0.02$) and 0.77 ($t[16]=3.05$, $p=0.004$) (see Table 4-25. Labor Leaders Experiment and Control Condition). The null of no learning could be rejected in a .10 and .05 one tailed test for the experimental group and in a .10, .05 and .01 one tailed test for the control group. It is appropriate to conclude that learning occurred under both conditions.

Table 4-23. Descriptive Statistics Pre-test to Post-test

Variable	Condition	N	Mean	SE Mean	StDev
True/False questions	Experimental	14	1.643	0.387	1.447
	Control	17	2.235	0.416	1.715
Labor Leader questions	Experimental	14	0.714	0.322	1.204
	Control	17	0.765	0.250	1.033

Table 4-24. T/F Experimental and Control Condition

N	Mean	StDev	SE Mean	95% Lower Bound	T	P (two tailed)
Experimental condition						
14	1.643	1.447	0.387	0.958	4.25	0.000
Control condition						
17	2.235	1.715	0.416	1.509	5.37	0.000

Table 4-25. Labor Leaders Experiment and Control Condition

N	Mean	StDev	SE Mean	95% Lower Bound	T	P (two tailed)
Experiment condition						
14	0.714	1.204	0.322	0.144	2.22	0.022
Control condition						
17	0.765	1.033	0.251	0.328	3.05	0.004

Did the Systems have Distinct Effects on Learning?

A second, more central issue to this study is whether the experimental and control condition had a meaningfully different effect on learning, which can be tested by using a difference in means tests for the T/F and Labor Leader questions. Consistent with the null hypothesis about the effect of the experimental and control systems on factual learning, the mean change in the control group score between the pre-test and the post-test is larger than the experimental group for both the T/F and Labor Leader questions. The difference on the True/False question is -0.59 items; it is -0.05 for the Labor Leader question. The negative signs are consistent with the hypotheses about the effect of the systems. The differences are not, however, close to being statistically significant. It is not possible to reject the null of no difference on the True/False question in a .05 or .10 test ($t[29] = -1.04$, $p = .15$) similarly for the Labor Leader question ($t[29] = -0.12$, $p = .45$).

Table 4-26. Differences in T/F and Labor Leader responses

Variable	Condition	N	Mean	SE Mean	St Dev
T/F questions	Experimental	14	1.643	0.387	1.447
	Control	17	2.235	0.416	1.715
Labor Leader questions	Experimental	14	0.714	0.322	1.204
	Control	17	0.765	0.250	1.033

Table 4-27. Differences in T/F responses*

Condition	N	Mean	StDev
Control	17	2.24	1.71
Experimental	14	1.64	1.45
Estimate for difference: 0.592			
T-Test of difference = 0 (vs >): T-Value = 1.04 P-Value = 0.153			

* Difference = μ (control) - μ (experimental)

Table 4-28. Differences in Labor Leader responses*

Condition	N	Mean	StDev
Control	17	0.76	1.03
Experimental	14	0.71	1.20
Estimate for difference: 0.050			
T-Test of difference = 0 (vs >): T-Value = 0.12 P-Value = 0.451 DF = 25			
* Difference = μ (control) - μ (experimental)			

Parallel results were obtained from the regression analyses. Each of the measures of change in knowledge was regressed for the experimental condition and the set of explanatory variables. In no case is the indicator for experimental condition close to being statistically significant. In the True/False model, the experimental condition indicator does not achieve significance in a .05 or .10 test ($\beta=-0.45$, $t=-0.86$, $p=.20$). Similarly for the Labor leader question ($\beta=-0.167$, $t=-0.36$, $p=.36$). Considering other variables, improvement on the T/F questions was negatively affected by respondents' total background score ($\beta =0.69$, $t=-3.65$, $p=0.001$) (i.e., the higher the initial score, the lower a participant's improvement from pre-test to post-test) and positively by total vocabulary score ($\beta=0.08$, $t=1.54$, $p=0.069$) (i.e., the higher the vocabulary score, the greater the score for Content Knowledge) (see Table 4-29. Regression of T/F Questions). The former result may be the result of two factors. First, it suggests those who had limited content knowledge prior to exposure learned more than those who had greater content knowledge prior to exposure and consequently learned less about the

Flint Sit-Down Strike. However, it may a partial ceiling effect. That is while no participant correctly answered 8 out of 8 T/F questions correctly, 11 people or 35% of the participants correctly answered 7 out of 8 questions correctly. No explanatory variables in the Labor Leader question had statistically significant coefficients (see Table 4-30. Regression of Labor Leaders Questions).

Table 4-29. Regression of T/F Questions

Predictor	Coef	SE Coef	T	P (two tailed)
Constant	4.691	1.578	2.97	0.007
Experimental Condition	-0.451	0.525	-0.86	0.399
Total Background Score	-0.691	0.189	-3.65	0.001
Total Vocabulary Correct	0.0820	0.053	1.54	0.137
Male	0.660	0.497	1.33	0.197
Library Catalog Use	-1.278	1.479	-0.86	0.396
Number of books checked out	-0.233	0.295	-0.79	0.438
Hours spent searching	0.027	0.234	0.12	0.909
R-Sq = 45.9% R-Sq(adj) = 29.5%				

* significant in a 10 percent test or better

Table 4-30. Regression of Labor Leaders Questions

Predictor	Coef	SE Coef	T	P (two tailed)
Constant	-0.764	1.377	-0.55	0.584
Experimental Condition	-0.166	0.458	-0.36	0.720
Total Background Score	-0.028	0.165	-0.17	0.867
Total Vocabulary Correct	-0.005	0.046	-0.10	0.920
Male	0.108	0.434	0.25	0.806
Library Catalog Use	1.998	1.290	1.55	0.135
Number of books checked out	-0.162	0.258	-0.63	0.536
Hours spent searching	0.074	0.204	0.36	0.720
R-Sq = 11.9% R-Sq(adj) = 0.0%				

* significant in a 10 percent test or better

Summary

To summarize, the pre-test, which measured participants' knowledge of labor history and labor leaders, found that participants did not vary systematically by condition. Next, the participants' attitudes toward the systems was reported. Testing revealed little evidence of a difference in participants' perception of the two systems by condition. Lastly, participant achievement on a vocabulary test was reported. Testing revealed that there were no statistically meaningful differences in the response to the vocabulary questions by condition.

Next, results for factual recall, synthesis and integration in which participants responded to a series of test questions with short answers and essays were reported. Consistent with the hypothesis, the control group scored significantly higher on recall than did the experimental group. However, contrary to the hypothesis the experimental group performed no better on the integrative essay than the control group. There was no statistically significant difference in either the difference in means or regression tests with respect to the number of themes or the number of facts though the control condition did make more connections. Similarly, for the tests for differences in the synthesis essay; there was not a statistically significant difference in the number of facts recalled between the two conditions in either the difference in means or regression test therefore the failure to find a difference was not due to variance. Finally, participants' retention was tested one week after their exposure to the system. Consistent with the hypothesis, participants in the control condition had greater factual recall than participants in the experimental condition. These results are summarized in Table 4-31, Summary of Difference by Condition. Differences in achievement that are significant in at least a .10 one tailed test are marked with an asterisk.

Table 4-31. Summary of Difference by Condition

Measure	Condition		
	Control μ	Experimental μ	Coef β
Recall	4.53	2.56	-2.0087*
Integration			
Themes	2.588	2.643	blank
Facts	3.294	3.571	-.001
Connections	2.64	2.88	-.464*
Overall score	8.765	8.857	-.421
Synthesis	3.64	3.85	.007
Attitude	4.93	4.95	.061
Retention	3.07	2.79	-.698*

From these results, it is concluded that the participants in the control condition were able to recall more facts in the test for recall than participants in the experimental condition. However, it cannot be concluded that more facts *in general* were remembered as there was no difference in achievement for the number of facts cited in either the test for factual integration or the test for synthesis. In addition, though participants in the control condition had higher achievement scores on the integration test, connections measure, there was no significant difference in overall performance between the two conditions. Lastly, differences in performance cannot be associated with the type of test. Though the tests for recall and retention were solely comprised of True/False and

multiple choice questions, the test for integration was an essay. In summary, there is little difference in the performance between participants using a linear hypermedia system and those participants using a context-rich hypermedia system and, where there were differences, the better performance was obtained from the control condition.

CHAPTER 5. ANALYSIS

How does this research relate to the hypermedia literature?

This research explores the effects of structural differences while controlling for individual differences and, as such, follows and builds on a dominant and promising theme in the hypermedia literature. In its use of CFT and parallels to Jacobson & Spiro (1995), it makes two contributions. First, it uses an underlying theoretical framework, which is important to developing research that can be generalized. Second, by replicating, in part, previously successful research, it adds to the CFT literature. While other hypermedia research, as has been noted, examines issues such as how people search for information, how does the presentation of information affect thinking and understanding and how do individual differences affect outcomes, no other research has explicitly contrasted interfaces to a simulated online library catalog and a simulated digital library. This research therefore extends the hypermedia research.

What does this research say about Cognitive Flexibility Theory?

CFT proposes seven principles that should guide instruction in ill-structured and complex domains (See Table 2-1. Cognitive Flexibility Principles). This research employed a sub-set of these principles, context-dependency and interconnectedness, which acknowledge and manage complexity. These principles were used because they can often be found in ADLIs and, as such, they are a realistic application. However, by employing a sub-set of principles, this research though it adds to the CFT literature does not address CFT in its totality.

This research also sought to replicate, in part, the findings of Jacobson & Spiro (1995), who used several principles from CFT including multiple conceptual representations of knowledge, the linking and tailoring of abstract concepts to different case examples and the introduction of domain complexity. Like Jacobson & Spiro, (1995), both studies concern historical events, both pre-tested undergraduate participants from large mid-western universities for domain knowledge and verbal comprehension and both studies used problem-solving measures in addition to factual tests and a test for retention. Again, however, by employing a sub-set of principles different from the earlier study, this research did not in fact replicate Jacobson & Spiro (1995). In addition, while both experiments employ principles from CFT, it isn't clear to what *degree* the various principles were involved. For example, while both experiments explore the interconnectedness of the themes, subtleties such as how apparent this interconnectedness was to participants or how effective the test questions were at exploring this are unknown. Further, though both sets of participants shared several sets of characteristics, it is impossible to know whether they were drawn from the same population. For example, the earlier study tested participants epistemic beliefs regarding the nature of learning and the structure of knowledge and found that those individuals who held more complex beliefs and were exposed to the experimental condition scored higher on problem-solving essays, The present study however did not test epistemic beliefs.

How does this research change our understanding of hypermedia systems?

A major contribution of this work is that it confirms prior research that suggests that, although the promise of hypermedia remains compelling, there is limited evidence for a measureable impact on learning (Dillon & Jobst, 2005; Wells; 2005; Dillon & Gabbard, 1998). However, a finding from this research and other research (Lee & Tedder, 2004; Eveland, Marton & Seo, 2004; Jacobson & Spiro, 1995) is that linear or indexed and/or drill-based designs can promote factual recall. What this means is that structural differences in hypermedia that follow a more minimal design, that parse nodes of information rather than presenting highly interconnected information aid learning.

For example, in the present research, the control condition (Figure 3-2. Linear/ control condition) attempted to replicate some of the functions and capabilities²⁸ of the Library of Congress' Online Catalog and Michigan State University's MAGIC online catalog. Specifically, individuals had to search for the information needed, locate a relevant bibliographic record (

Figure 3-3. Linear/ control bibliographic record) and then click within the object to access the actual information. Each bibliographic record provided the type of information (e.g., text, audio and images), title, author, description, date and subject as well as a direct link to the information itself. Either this process or stepped access to information or some combination may have helped participants in the control condition to have greater recall than those in the experimental condition.

Though this research has limitations, (see the section entitled "Limitations" in Chapter 1 and the section entitled "Alternative conclusions" in this chapter for a

²⁸ Current as of March 2006.

discussion of this issue), it also has important strengths some of which have been discussed (see also the section entitled “Why this research” in Chapter 2 for a discussion). However, one, the research design, still requires discussion. The research randomly assigned 33 participants to one of two conditions. All participants were pre-tested for knowledge of the Flint Sit-Down Strike and vocabulary knowledge in order to ensure that the groups were equivalent for testing purposes. Later, additional characteristics including use of a library catalog, amount of material they had previously checked out of a library and time spent searching the internet were also gathered again to ensure that the groups were equivalent. Both learning environments underwent usability testing and then had issues mitigated. Prior to actual data collection, pilot testing was conducted to ensure that the testing itself would be focused on the experiment itself rather than juggling process with data collection. The measures used were developed and informed with expertise from domain experts. Lastly, participants recall, synthesis and integration was measured immediately following exposure and one week after testing, their retention was measured. In summary, the experiment was well formulated and carried out systematically and is a significant contribution as an overall methodology for comparing and testing interfaces

The reason why summarizing this strength is helpful is because of what this research *did* find. Participants in *both* conditions increased their knowledge of the Flint Sit-Down strike between pre-test and post-test because of exposure to their respective learning environments. In addition, participants in the simulated library catalog condition were able to recall 1.97 more items than the experimental

group. Lastly, while ADLIs have captured the imaginations of many, the participants in each condition reported high levels of satisfaction with their system. This is important because tens of thousands of libraries provide library catalogs for their patrons to conduct research and, as such, have a large installed base. Additionally, there is an assumption that ADLIs are more satisfying than online catalogs for patrons to use. This has three implications. First, though library catalogs were initially conceptualized as “finding–catalogues” (Lubetsky, 1969), patrons are learning about the content of the resources themselves, how they interconnect with other resources and inform their research question. From a learning perspective, ADLIs may have similar cognitive affordances to online library catalogs though this conclusion is provisional as both the ADLI and simulated online library catalog contained a limited number of multimedia resources on a highly specific topic and was tested with a very specific population.

Second, it is important to understand both the benefits and limitations of ADLIs. Though this study only examines a small aspect of this broader issue specifically interface design, it is important given the sizable amount of funding being spent on ADLIs, including their development, maintenance and on new applications. At the same time that significant efforts are being spent on ADLIs, traditional cataloging budgets and staff are being reduced. This research does not seek to determine whether this ongoing trend is appropriate or not. However, it is important from conceptual perspective to understand that patrons may be

learning even as they search for materials in the online library catalog so decreasing cataloging efforts may affect their learning.

Third, it is important to consider the relative costs of both online library catalogs and ADLIs in the context of their contributions. What do each cost to develop and maintain relative to their function as learning environments? Further, given limited resources, should the development of each continue as it has or, for example, should there be a greater division in their purpose?

Alternative conclusions

There are many alternative conclusions to the research findings that need to be considered. The first set of alternative conclusions relates to the study's theoretical framework. First, it may be that the design did not adequately reflect CFT theory. For example, the CFT principles chosen, context-dependency and interconnectedness, may have been insufficient as manifest in the system. Alternatively, the use of other principles or more principles would have yielded different results. That is it was the choice of the two principles chosen and their incorporation rather than CFT itself that influenced the results. Second, it may be that a maximum exposure of 2 hours to the materials was insufficient to observe the effects of CFT. That is, if both sets of participants had had a longer exposure (perhaps to a larger collection of resources), then the results for the experimental condition would have been significantly different. For example, it may have been that participants in the experimental condition would have been willing to dig deeper and delve longer than participants in the control condition. Third, the size of the collection itself may have influenced results. That is, if the collection

contained 146 resources and not only 73 resources, different and advantageous results for the experimental condition would have been obtained. Again, it may have been participants in the experimental condition would have been more motivated to learn and consequently have learned more.

The second set of alternative conclusions relates to the study's selection and pre-testing. First, it may be that a random sampling of the population might have yielded different results rather than the sample of self-selected, self-declared history majors at Michigan State University that was used. Second, it may be that different tests would have provided different results. For example, while participants were pre-tested for verbal comprehension and this was used as a general measure of intellectual ability, tests that assessed participants' actual writing ability, interest in the subject and/or motivation might have demonstrated systematic differences between the participants themselves and it was these differences that influenced the findings rather than the condition. Third, though the research attempted to control for intervening variables such as background knowledge, comprehension, etc. participants were not asked about physical limitations such as vision or hearing impairments. This means that differences in abilities to receive or even process the information that they were exposed to were not measured.

The third set of alternative conclusions relates to the study's treatment conditions. First, it may be that the participants in the linear, control condition performed better than anticipated because they had to construct an understanding of the information. That is, that the process of having to search for

discrete resources and not having the contextualized information available may have meant that participants in the control condition had to mentally construct coherent understandings of how the resources related to each other and it was this process that ultimately enhanced their ability to recall, synthesize, integrate and retain information. Second, it may be that immediate access was a key determinant in how participants performed. That is, immediate access to the electronic resources enabled both groups to recall, synthesize, integrate and subsequently retain the information in general rather than assignment to condition. Third, it may be that the nature of the resources themselves – audio, images and text – influenced results in a way that may have been different were all the resources text or even audio. For example it may be that college-aged students are more comfortable listening to audio than an older age group might be. Fifth, it may be that the use of general resources (e.g., literature, biology, etc.) or even resources on strikes in general rather than only the Flint Sit-down resources would have produced different results. That is, it may be that some participants may have been less or even more motivated and that these differing levels of motivation subsequently influenced results.

The fourth set of alternative conclusions is more general. First, the design of either or both the control or experimental system may have produced results in a manner not explored herein. That is, though each system was realistic and went through a heuristic evaluation; perhaps the design or implementation was inadequate and ultimately influenced the results. Secondly, it may be that the results obtained were due to some degree by the participants' familiarity with

online library catalogs or ADLIs. For example, differences may have been found if one set had been less familiar with their assigned condition.

These alternatives are presented in part to emphasize the difficulty of designing and implementing studies that compare interfaces. For example, this study went to great lengths to equalize aspects of the two conditions so that neither was disadvantaged by the short time span of the study. That decision, however, may have neutralized some of the advantages of the experimental condition, being able to reach content faster. Many other possibilities are suggested by the alternative hypotheses mentioned above, pointing to the complexity of establishing equivalent learning conditions for a head-to-head comparison of interfaces. This problem as demonstrated in the Literature Review and elsewhere has occurred repeatedly in studies of the impact of technology on learning.

Recommendations for further research

Many recommendations for future research can be directly inferred from the section entitled "Alternative conclusions." These recommendations are of two different types. The first type involves basic changes to the research design such as the random selection of a sample or increasing the number of resources and the amount of time participants are exposure to and then testing for differences in outcomes. The second type involve larger changes to the research such as selecting different principles from CFT or exposing participants to different domains or resources altogether. However, further specific recommendations for future research based on the existing hypermedia literature and the present

research are important to consider and, as such, several follow. These recommendations in conjunction with the Literature Review and Alternative Conclusions represent a significant contribution as they can be used to guide future research.

A logical follow-up study might include varying the nature of the electronic resources in terms of the type and number available to participants. More specifically, a study that involved the use of American Memory, which embodies several CFT principles, and the Library of Congress's (LOC) online catalog and that allowed participants to choose from one of two historical topics would be significant. American Memory and the LOC online catalog are widely used information resources, that contain wide-ranging and in many cases overlapping multimedia content and, as such, represent ecologically valid systems. Though no experiment is without limitations, this would address two limitation of the present research including the nature of the resources themselves and the fact that these systems contained a much lower number of resources than one would typically find in the real world. Allowing participants to choose between one of two topics while pre-testing for prior knowledge, vocabulary comprehension, prior use of systems including the internet, attitudes and gender, would ideally allow participants to choose a topic of interest and make the task itself more engaging. Participants might be tested one of two ways, as participants were in the current research or they might be given a series of issues to explore and then subsequently answer questions about.

Alternatively, a study that allowed participants to choose from one of two topics focused on a physical science might yield interesting results. Such a study would mirror the one outlined above. However, by focusing on a scientific domain it would begin to explore the potential effect of the subject matter on outcomes. That is, are there differences in learning when participants study astronomy or biology versus topics from the social sciences?

Research might also focus on prescriptive artifact design, a conclusion also mentioned by Jacobson and Spiro (1995). This research would enumerate and define specific characteristics that CFT systems embody. The goal of this effort would be to enable future CFT research that is based on specific standards and designs that can be replicated study to study. What this would mean is that instead of noting the presence or absence of CFT principles, the importance or magnitude of the principle and subsequently its impact on results might be better explored. This research would take into account learning goals (e.g., factual knowledge acquisition, problem solving, information finding, etc.) and population characteristics (e.g., prior knowledge, cognitive/learning styles, etc.).

Lastly and more broadly, in addition to CFT, there are several hypermedia theories in the literature that warrant systematic testing dual coding (Pavio, 1986), Mayer and colleagues' generative theory of multimedia learning (Mayer, 2001) and related theories such as cognitive load theory (Sweller, 1988). These theories are not necessarily antithetical to one another and may inform hypermedia research and prescriptive design.

APPENDICES

Appendix A: Heuristic Evaluation - A System Checklist

Please use the comments column for clarifications or additional thoughts, questions, etc.

You may also use the space below each table or on the back of each page.

1. Visibility of System Status

The system should always keep user informed about what is going on, through appropriate feedback within reasonable time.

#	Review Checklist	Yes	No	N/A	Severity Rating*	Comments
1.1	Does every page display begin with a title or header that describes screen contents?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
1.2	Is there some form of system feedback for every user action?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
1.3	After the user completes an action (or group of actions), does the feedback indicate that the next action can be started?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
1.4	Is there visual feedback about which choices are selectable?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
1.5	Are there are observable delays (greater than ten seconds) in the system's response time?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
1.6	Are response times appropriate?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		

Table 1 (cont'd).

1.7	Is continuity of thinking required by needing to remember information through several screens?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
1.8	Are high levels of concentration necessary for remembering information from screen to screen?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
1.9	If users must navigate between multiple screens, does the system provide navigational aids?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
1.10	Is it relatively easy for the user to understand where they are in the system?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
1.11	Is it relatively easy for the user to understand where they might wish to go next in the system?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		

* Severity Rating – please provide a rating for *only* those items you have checked as “no” in the previous column.

0 I don’t agree that this is a usability problem at all.

1 Cosmetic problem only – need not be fixed unless extra time is available on project

2 Minor usability problem – fixing this should be given low priority

3 Major usability problem – important to fix, so this should be given high priority

4 Usability catastrophe – imperative to fix this before product can be released

2. Match between System and the Real World

The system should speak the user's language, with words, phrases and concepts

familiar to the user, rather than system-oriented terms. The system should follow real-

world conventions, making information appear in a natural and logical order.

#	Review Checklist	Yes	No	N/A	Severity Rating*	Comments
2.1	Are menu choices ordered in the most logical way, given the item names, and the task variables?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
2.2	If there is a natural sequence to menu choices, has it been used?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
2.3	Does related and interdependent information appear on the same screen?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
2.4	On data entry screens, are tasks described in terminology familiar to users?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
2.5	Are field-level prompts provided for data entry screens?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
2.6	Is the labeling and language clear throughout the site?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		

Table 2 (cont'd).

2.7	Is the labeling and language clear for each record?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
-----	--	---	--	--

* Severity Rating – please provide a rating for *only* those items you have checked as “no” in the previous column.

- 0 I don’t agree that this is a usability problem at all.
- 1 Cosmetic problem only – need not be fixed unless extra time is available on project
- 2 Minor usability problem – fixing this should be given low priority
- 3 Major usability problem – important to fix, so this should be given high priority
- 4 Usability catastrophe – imperative to fix this before product can be released

3. User Control and Freedom

Users should be free to select and sequence tasks (when appropriate), rather than having the system do this for them. Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Users should make their own decisions (with clear information) regarding the costs of exiting current work. The system should support undo and redo.

#	Review Checklist	Yes	No	N/A	Severity Rating*	Comments
3.1	Can users cancel out of operations in progress?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
3.2	Are character edits allowed in data entry fields?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
3.3	Do users have the option of either clicking on menu items or using a keyboard shortcut?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
3.4	If the system has multiple menu levels, is there a mechanism that allows users to go back to previous menus?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
3.5	Can users move between levels and topics easily?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		

* Severity Rating – please provide a rating for *only* those items you have checked as "no" in the previous column.

0 I don't agree that this is a usability problem at all.

- 1 **Cosmetic problem only – need not be fixed unless extra time is available on project**
- 2 **Minor usability problem – fixing this should be given low priority**
- 3 **Major usability problem – important to fix, so this should be given high priority**
- 4 **Usability catastrophe – imperative to fix this before product can be released**

4. Consistency and Standards

Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

#	Review Checklist	Yes	No	N/A	Severity Rating*	Comments
4.1	Are buttons adequately labeled?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
4.2	Does each window have a title?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
4.3	Is vertical scrolling possible, where appropriate?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
4.4	Does the menu structure match the task structure?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
4.5	If "home" is a choice, is it easy to locate and does it appear consistently?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
4.6	Are menu titles used consistently and consistently placed?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
4.7	Are field labels and fields distinguished typographically?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
4.8	Are field labels consistent from one data entry screen to another?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		

Table 4 (cont'd).

4.9	Are attention-getting techniques used with care?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
4.10	Are there no more than four to seven colors, and are they far apart along the visible spectrum?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
4.11	Is a legend provided if color codes are numerous or not obvious in meaning?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
4.12	Is the most important information placed at the beginning of the prompt?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
4.13	Does the structure of menu choice names match their corresponding menu titles?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
4.14	Are commands used the same way, and do they mean the same thing, in all parts of the system?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		

* Severity Rating – please provide a rating for *only* those items you have checked as

“no” in the previous column.

0 I don't agree that this is a usability problem at all.

1 Cosmetic problem only – need not be fixed unless extra time is available on project

- 2 Minor usability problem – fixing this should be given low priority
- 3 Major usability problem – important to fix, so this should be given high priority
- 4 Usability catastrophe – imperative to fix this before product can be released

5. Help Users Recognize, Diagnose, and Recover From Errors, and 6. Error Prevention

Error messages should be expressed in plain language.

Even better than good error messages is a careful design which prevents a problem from occurring in the first place.

#	Review Checklist	Yes	No	N/A	Severity Rating*	Comments
5.1	Do prompts imply that the user is in control?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
5.2	Are error messages worded so that the system, not the user, takes the blame?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
5.3	Are error messages grammatically correct?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
5.4	Do error messages avoid the use of exclamation points?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
5.5	Do error messages suggest the cause of the problem?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
5.6	Do error messages make sense?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
5.7	Do error messages clearly indicate what action the user needs to take to correct the error?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
6.1	Are menu choices logical and distinctive?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		

Table 5 (cont'd).

6.2	Are images, text and audio files/sounds easy to access?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
-----	---	-----------------------	-----------------------	-----------------------	--	--

* Severity Rating – please provide a rating for *only* those items you have checked as “no” in the previous column.

- 0 I don’t agree that this is a usability problem at all.
- 1 Cosmetic problem only – need not be fixed unless extra time is available on project
- 2 Minor usability problem – fixing this should be given low priority
- 3 Major usability problem – important to fix, so this should be given high priority
- 4 Usability catastrophe – imperative to fix this before product can be released

7. Make objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible whenever appropriate.

#	Review Checklist	Yes	No	N/A	Severity Rating*	Comments
7.1	Are all data a user needs on display at each step in a transaction sequence?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
7.2	Are prompts, cues, and messages placed where the eye is likely to be looking on the screen?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
7.3	Does the system gray out or delete labels of currently inactive options?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
7.4	Is white space used to create symmetry and lead the eye in the appropriate direction?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
7.5	Have items been grouped into logical zones, and have headings been used to distinguish between zones?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		

Table 7 (cont'd).

7.6	Have zones been separated by spaces, lines, color, letters, bold titles, rules lines, or shaded areas?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
7.7	Are zones used to break long input strings into "chunks"?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
7.8	Is color highlighting used to get the user's attention?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
7.9	Is color highlighting used to indicate that an item has been selected?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
7.10	Are borders used to identify meaningful groups?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
7.11	Has the same color been used to group related elements?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
7.12	Is color coding consistent throughout the system?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
7.13	Is color used in conjunction with some other redundant cue?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
7.14	Is there good color and brightness contrast between image and background colors?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		

Table 7 (cont'd).

7.15	Has color been used with discretion?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
7.16	Has color been used specifically to draw attention, communicate organization, indicate status changes, and establish relationships?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		

* Severity Rating – please provide a rating for *only* those items you have checked as “no” in the previous column.

- 0 I don’t agree that this is a usability problem at all.
- 1 Cosmetic problem only – need not be fixed unless extra time is available on project
- 2 Minor usability problem – fixing this should be given low priority
- 3 Major usability problem – important to fix, so this should be given high priority
- 4 Usability catastrophe – imperative to fix this before product can be released

8. Flexibility and Minimalist Design, and 9. Aesthetics

Accelerators-unseen by the novice user-may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions. Provide alternative means of access and operation for users who differ from the "average" user (e.g., physical or cognitive ability, culture, language, etc.)

Dialogues should not contain information, which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

#	Review Checklist	Yes	No	N/A	Severity Rating*	Comments
8.1	On data entry screens, do users have the option of either clicking directly on a field or using the 'Enter' key as a shortcut?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
8.2	Is there redundant and distracting information used?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
8.3	Is there missing information or explanations?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
9.1	Is only information that is essential to decision making displayed on the screen?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		

Table 8 & 9 (cont'd).

9.2	Is all information that is essential to decision making displayed on the screen?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
9.3	Are all navigation text boxes, images, and text which are in a set visually and conceptually distinct?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
9.4	Are meaningful groups of items separated by white space?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
9.5	Does each data entry screen have a short, simple, clear, distinctive title?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
9.6	Are field labels brief, familiar, and descriptive?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		
9.7	Are menu titles brief, yet long enough to communicate?	<input type="radio"/> <input type="radio"/> <input type="radio"/>		

* Severity Rating – please provide a rating for *only* those items you have checked as “no” in the previous column.

0 I don’t agree that this is a usability problem at all.

1 Cosmetic problem only – need not be fixed unless extra time is available on project

2 Minor usability problem – fixing this should be given low priority

- 3 Major usability problem – important to fix, so this should be given high priority
- 4 Usability catastrophe – imperative to fix this before product can be released

Heuristic Evaluation

A System Checklist

Primary Source

Making Computers-People Literate. © Copyright 1993.

By

Elaine Weiss

ISBN: 0-471-01877-5

Secondary Source

Usability Inspection Methods. © Copyright 1994.

By

Jakob Nielsen and Robert Mack

ISBN: 1-55542-622-0

Appendix B: Content knowledge Flint Sit-Down Strike

1) The Flint Sit-Down Strike took place in:

Michigan in the 1930s

Michigan in the 1950s

Michigan in the 1970s

2) The Flint Sit-Down Strike concerned which Michigan company:

Chrysler

Ford

General Motors

3) The decision to actually begin the strike was made by a majority of unionized workers.

True

False

4) The major factors that contributed to the strike were:

Reduced retirement packages

Wages and working conditions

Reduced health care benefits

5) The strikers used a new approach to striking one wherein workers:

Marched to the mayor's home to announce the strike's start

Occupied the facility

Engaged in a work slow-down

6) The company's stance toward unionization was:

One of partnership

One of caution

One of opposition

7) The National Guard was called up during the strike to prevent both parties from harming one another.

True

False

8) Which five of the following people gained fame as national labor leaders?

- Mary Harris Jones

- William Randolph Hearst

- James McNelly

- Eduardo Montiel

- Eugene Debs

- Arthur "Bud" Fletcher

- Anita Elder Green

- Kirk P. Hogan
- George Meany
- James P. Hoffa
- Jerry Lewis
- John Wayne
- Michelle McNelis-Valvano
- G.W. Singer
- John L. Lewis

Appendix C: Evaluation Measures

Instructions: Answer sections I, II and III to the best of your ability. Do NOT refer to library website.

I. For six of the resources (audio, text and/or image) in the system, write down as much as you recall about the resource such as its title, author and/or subject.

1.
2.
3.
4.
5.
6.

II. Answer the following question by writing one brief essay of at least 250 words.

Keeping in mind working conditions, strike methodology, conditions during the strike, and community responses, how did General Motors try to control its workers?

III. Answer the following question by writing one brief essay of at least 250 words.

Using specific examples, what were people's rationales in their stance for or against the strike?

Appendix D: Attitude Survey

Instructions: Answer the questions based on your own feelings.

	QUESTION	SCORE					
		Strongly agree	Agree	Weakly agree	Weakly disagree	Disagree	Strongly disagree
1	I liked the Flint Sit-Down Strike system.	Strongly agree	Agree	Weakly agree	Weakly disagree	Disagree	Strongly disagree
2	I would recommend this system to other students.	Strongly agree	Agree	Weakly agree	Weakly disagree	Disagree	Strongly disagree
3	This system was easy.	Strongly agree	Agree	Weakly agree	Weakly disagree	Disagree	Strongly disagree
4	I learned a lot about the Flint Sit-Down Strike using this system.	Strongly agree	Agree	Weakly agree	Weakly disagree	Disagree	Strongly disagree
5	It was easy to understand where I was in the system.	Strongly agree	Agree	Weakly agree	Weakly disagree	Disagree	Strongly disagree
6	It was easy to find information in the system.	Strongly agree	Agree	Weakly agree	Weakly disagree	Disagree	Strongly disagree
7	It was easy to move around in the system.	Strongly agree	Agree	Weakly agree	Weakly disagree	Disagree	Strongly disagree
8	The best part of the program was...						
9	The worst part of the program was...						

Appendix E: Example of the Integrative, Synthetic and Retention Essays

Exemplar Integration Essays

The following two essays, which each received *high scores*, serve as exemplar individual responses to essay number 1, the integration essay:

Keeping in mind (1) working conditions (WC), (2) strike methodology (SM), (3) conditions during the strike (CDS), and (4) community responses (CR), how did General Motors try to control its workers?

Each of the two essays includes multiple themes, facts and connections with the result that they received high scores. The first essay received a score of 12 (4, 4 and 4) and the second essay received a score of 11 (3, 4 and 4). Themes have been highlighted in gray. Following each discrete fact is a number in superscript that is a running tally of the number of facts in each essay. As both essays contained more than two themes and more than for facts that were clearly connected within a coherently structured essay, they each received a top score of 4 on the connections measure.

Participant essay:

General Motors attempted to control its workers largely through coercion, harassment, and propaganda. Before the strike even began, the normal tactic for worker control was the treat of a lost job¹. This would have been taking place during The Great Depression when jobs were scarce, and as one man said in his interview, there were 5000-1 men waiting outside the plant to take his job². As

such, if workers wanted to keep their job, they had to put up with whatever the plant wanted³. In addition, GM company men would keep a profile on each of its workers⁴, so if a man with a family⁵ and a house complained⁶, they would know he could not afford to lose his job. Strike methodology reflected this fear of losing a job; rather than walk out and let scabs take their jobs, men sat in the factory⁷, prevent production⁸ and the loss of jobs⁹ to other men. Harassing of families was another tactic employed by GM 'goons.'¹⁰ Men would go to the homes of union workers, bang on the doors and threaten violence and murder on the wives of union men. This, they thought, would get the women to try and persuade their husbands to come home¹¹. Worse off though were the men in the plants. As GM largely controlled the city of Flint, they could have the police¹² in riot gear ready to shoot or tear gas the men. Also, they had judges¹³ who owned stock in GM give the ruling ordering the men out of the plant. Lastly, GM sent out propaganda biased against the workers. Schoolchildren¹⁴ had to write on why the strike was wrong, and newspapers portrayed the men as 'reds' and 'radicals.'¹⁵ Thinking this would turn public opinion against the union, GM sent out as much biased news as they could.

Participant essay:

Before the strike, GM controlled its workers through the company union¹. The company union was not a labor union in the sense that the union which organized the strike was. Its primary purpose was to gather information² about the workers in order to keep them from agitating against the poor working

conditions³ or for higher wages⁴. By spying on their employees, they were able to blackmail them into putting up with low wages and poor working conditions. For example, they would find information about a worker's family and his financial situation⁵, and then say to him, if he tried to oppose them, 'But you have a wife and six children, you can't really afford to lose your job right now.' GM also controlled public opinion⁶ in Flint because the local newspaper⁷ was controlled by GM. When the strike appeared to be imminent, GM tried to exercise its power over the workers by moving the die presses⁸ from Fisher 2 to another plant. By doing this, GM could have ensured that they were still able to produce cars even after the workers occupied the plant⁹. This failed, however, because the union leaders heard about this and decided to start the strike early¹⁰ than they had planned in order to stop GM from moving the dies. During the strike, GM tried to regain its hold over the workers by cutting off the heat¹¹ to the plant (the strike took place in winter) and by asking the governor¹² to order the National Guard to remove the strikers from the plants. Their first tactic failed because the workers started burning burlap¹³ in order to keep warm. Because burlap was very expensive, GM relented quickly and delivered coal¹⁴ to the plant. Their appeals to the governor also failed because the governor was sympathetic¹⁵ to the labor movement.

Each of the two essays includes minimal themes, facts and connections with the result that they received *low scores*. The first essay received a score of 5 (2, 2 and 1) and the second essay received a score of 6 (3, 1 and 2). Themes have been highlighted in gray. Following each discrete fact is a number in superscript that is a running tally of the number of facts in each essay. As both essays contained a lesser number of themes and facts that were poorly or loosely connected within poorly structures essays, they each received lower scores on the connection measure.

Participant essay:

Before the strike even began General Motor's had a strong lock over many of its employee's. One thing they did was pay different wages for the same job¹, giving special treatment to some employees. Another, was to treat workers like they were simply a body on a line, and therefore they could be replaced. This allowed for harsh treatment by foreman and other bosses. Workers were not often allowed to take a bathroom break², get a drink of water³, eat lunch⁴, or leave if they felt ill⁵. One employee once fell to the floor while working on the assembly line, he had previously asked to go outside and get some fresh air and relief from the hot environment, but had been refused. Laborers found themselves working for years at a time and never receiving wage, often they would receive pay cuts with rarely a reason why⁻¹. The GM company plant reigned with tyranny over its Auto employees. When researching the above

treatment and conditions it seems that it would come as no surprise when workers began the infamous 'sit down' strike. General Motors was shocked and confused² when laborers in Flint's automotive plant held a sit-down strike only days after Christmas. Although they were unaware of the strike in general³, the most shocking aspect was the type of strike used by the workers. A 'sit down' strike, which had only come about in popularity was now being held at their own plant. The response was immediate, General Motors informed the police, however there was little the police⁶ could do seeing as the National Guard was soliciting protection⁴ outside the peaceful inside sit down strike. They also at one point snuck people in through tunnels that ended in a bloody battle, Bull's Run⁷, as it has been called involved workers and laborer's fighting tooth and nail with clubs⁸ and other items against each other. The result was a few broken bones and the retreat of the owners, not the employees already inside. The sit down strike was a success for many reasons; it showed a want for resolution. The workers stayed inside the building⁹, and therefore seemed to convey a desire to keep on at the plant if things could improve. It also showed the company that they could protest in a peaceful manner and would not use violence gain it¹⁰. The method of the strike was important because it got numerous groups in the community behind the workers.

Participant essay:

General Motors sent in 'Goons'¹ to help control the strike. They had the goons come in and try to get the strikers to leave and do their job so the plant could get on. The goons were the equivalent to mobsters. In fact some of the goons actually were mobsters who wanted to go legit or just wanted to appear as they were. So as you can imagine these were some pretty tough ruthless guys. General Motors blamed the outside influences as being 'red' and communists². They tried to remove the dies³ to have the car's be made else where for example in Atlanta⁴, Georgia. The corporation also instituted propaganda about the strikers and the head of the union being communists and reds. This made a lot of the strikers angry because they were American and felt that they were completely so and did not wish to have their patriotism lessened by this. General Motors also felt as though the public was on their side. The people of Flint did not like the company being on strike⁻¹. It was bad for the community General Motors of course played this out as much as they could. General Motors tried to down-play how bad the strike was and how terrible the working conditions were to the stockholders⁻². Even though it was very difficult if not impossible for General Motors to get any work done in the plant and get any products out on the market. Overall General Motors did what any corporation would do and that is use any tactic available to get things moving along as usual.

Exemplar Synthesis Essays

The following two essays, which each received *high scores*, serve as exemplar individual responses to essay number 2, the synthesis essay:

Using specific examples, what were people's
rationales in their stance for or against the strike?

Each of the two essays includes multiple facts with the results that they received the highest score of 6. Following each discrete fact is a number in superscript that is a running tally of the number of facts in each essay.

Participant essay:

The automotive workers went on strike for two main reasons: wages¹ and working conditions². In general, the wages of the workers had been rising¹, but there were major discrepancies in pay³. An assembly line worker might be making about 46 cents, while the assembly line worker right next to him was making 60 cents or more⁴. With General Motors worth more than a billion dollars, the workers felt they should all be moved toward the higher end of the pay scale². The working conditions, though, were the biggest problem for the workers. The factories were hot⁵, the workers could not take breaks for necessities such as water⁶, restroom use⁷ was strictly controlled, and the assembly line⁸ moved so fast that the workers felt they could not do quality work. One automotive worker said working in the factory felt like being treated 'like a dog.'⁹ Other minor issues led workers to strike. One such issue was seniority¹⁰. There was no system of seniority in the automotive factory. A worker could be laid off whenever the

factory needed to cut costs¹¹. Some workers felt seniority should be taken into account so that people were 'laid off when it was your turn and rehired when it was your turn.' The bonus system¹² also caused resentment among the workers. Workers were not given raises based on how long they had been employed or how well they did their work¹³. At some plants, a worker made 50 cents¹⁴ from the time he was hired until the time he retired. Another plant had a bonus system where workers were given bonus money instead of raises. However, for workers in a plant running on the bonus system, a cut in pay was devastating because a worker might only be making 44 cents to begin with. Most people opposed the strike because their economic or political interests were threatened. The strike put a halt to most economic activity in Flint¹⁵, so the local businesses suffered. The workers stopped buying from the local shops because they were no longer earning money in the factories. The businessmen of the city wanted the strike to end because their livelihood was threatened. The biggest reason for opposing the strike, however, was that it was illegal¹⁶. Even some of the workers admitted that the methods of the strike violated the U.S. Constitution. The workers had seized the property of the company¹⁷, and the Constitution guarantees the company's right to its property.

Participant essay:

The majority of those who were for the strike were workers or union organizers¹. The rationale of the workers is the most straightforward. The conditions² in the GM plants in Flint were very poor. Workers often got sick and

fainted due to the heat³ and the lack of water⁴ in addition to the long hours⁵ that they worked. Not only were they forced to work in very poor conditions, but they also received very low wages⁶. Some of the workers were satisfied with their wages, but even a number of those workers⁷ chose to join the union because they believed in equality and wanted their fellow workers to make as much money as they did. These workers were also motivated by the poor working conditions, which affected them all equally. The catalyst for this strike was the speed-up⁸ that GM instituted in order to raise production levels. The already poor working conditions were worsened by the strike¹ and the amount of work that they were required to do left them exhausted. This was the 'last straw' for many workers, and it motivated them to join the union. The organizers of the strike, including Mortimer, Travis⁹, and the Reuther brothers¹⁰ and community leaders who supported the strike, including Genora Johnson¹¹ were motivated by their political beliefs. Although communism and socialism did not play an explicit role in the organization of the strike¹², it was later revealed that these leaders had leftist political leanings and some of them were socialists¹³. Because of these political convictions, and in Johnson's case because of a personal interest in the strike, they supported the workers. Those in the community who opposed the strike did so for several different reasons. Private individuals and GM shareholders opposed the strike because they felt that a sit-down strike was in violation of the American value of property rights¹⁴. Some local churches opposed the strike because of their commitment to non-violence¹⁵. Others in the community opposed the strike for more practical reasons. Local business owners

were against the sit-down strike because it stopped the flow of business and money into their shops¹⁶. GM officials and others in Flint opposed the strike because it stopped production¹⁷.

The following two essays, which each received *low scores* of 2, serve as exemplar individual responses to essay number 2, the synthesis essay:

Participant essay:

People's rationales for supporting the strike are mainly due to their close relationship¹ with the conditions in the factories, either as a worker or the wife of a worker. One man described life inside the factory as a fire that was never put out². The workers were expected to produce at outrageous rates. One man describes a neighbor of his as working through his lunch hour in order to keep up the pace in order to keep his job³. These conditions took a serious toll on all of the workers. Flint workers were described as being grayish in color⁴, almost as if they were victims of tuberculosis. Wives were extremely worried about their husbands' health because of this. Also, conditions in the factories were as such that during an extremely hot summer, the deaths within the factories numbered in the hundreds⁻¹. Breaks would have been impossible, since that would only put a person further behind in production⁵. When people did not support the strike, they also did so because it hurt them. However, for the most part, it hurt these people only financially. Local businesses lost money due to the strike⁶. General Motors lost money due to lost production time⁷, and as a result, their stockholders lost money. One man even describes not caring about the working

conditions in the factories. He just wanted his business to make money².

Everyone had a personal stake in the Flint strikes. However, in the case of the supporters, it was a vested interest in their own health and well-being, as well as financial. Those who did not support the strike did not because they lost money during it³. In the case of the General Motors stockholders, these people actually made money due to the practices of General Motors in the first place.

Participant essay:

There were many rationales of people during the strike. While a strike is usually given the automatic assumption of being about the workers, there is a cyclical affect that a strike can have on everything involving it. Workers strike and lose money, the plant loses money etc. However, the businesses in the community lose money as well and can possibly lose sustainability because their consumers are no longer consuming goods and services¹. Thus, those stores can begin to have a surplus of goods that they cannot even rid themselves of with a sale. A lack of capital can also lead to other things such as a decline in health and things like that. With all of that said, the store owners were not receptive to those in the strike, shunning them away². However, there were some instances of those who had worked at the factories, such as the farmer³ who worked at a plant and expressed his sentiments. His sentiments turned into condolences, of which was a whole hog⁴ to feed the strikers. The employers wanted to end the strike quickly in order to begin to make more money. However, without the strike, it can be argued that the employers would have given very little

thought to the condition of their exploited workers. In fact, until the money that they hoarded for themselves began to dwindle, they would care nothing for their workers. It takes agitation in order to get results from an exploiter. The employers cared nothing about the rationale of the workers until the worker began to want equality. The rationale of the exploiter did not foresee the worker eventually wanting equality. This is ignorance on behalf of the GM employers.

Exemplar Retention Essays

The following two essays, which each received *high scores*, serve as exemplar individual responses to essay number 3, the retention essay:

Imagine that you are an employee in one of G.M.'s Flint plants. A fellow worker approaches you and asks you to participate in the sit-down strike that has just broken out in Fisher One and Two. Would you join in the strike? Using specific examples, name three reasons why you would or why you would not?

Each of the two essays includes multiple facts with the results that they received the highest score of 4. Following each discrete fact is a number in superscript that is a running tally of the number of facts in each essay.

Participant essay:

I would join the strike. For one, I am assuming that my working conditions would have been abominable¹; certain workers cited having to work in extreme heat past to the point of passing out², lack of replacement workers³ to allow one a water break, and the lack of time for eating⁴. In addition, Workers cited mistreatment by management; they were asked about person information that could be used against them later⁵, and there was little to no room to advance in wages⁶; you were paid what they paid you. As a last point, I know I would have the support of the newly elected governor, who was a New Deal politician⁷, and sympathetic to the plight of unions.

Participant essay:

If I were approached to participate in the sit-down strike, I believe I would agree to participate. The conditions in the plant were such that I would probably be in relative poor health, due to working with few breaks¹, and needing to produce² very high amounts of product. There would be no other way to change the working conditions than to participate in the strike. I would also participate because the sit-down strike would most likely be less violent³ than other forms of striking. Since we would all already be inside the plant, there would be no large mobs of people trying to fight their way inside, or trying to fight for someone to hear them. Finally, I would participate because the sit-down strike is a good way to guarantee that management would have to listen. Production has to stop in a sit-down strike⁴; this is why it was very important that they began the sit-down strike prior to GM being able to remove the dies⁵. GM had to listen to its strikers, because it was losing valuable production time⁶ and money with people just sitting in its factories, not working.

The following two essays, which each received *low scores* of 1, serve as exemplar individual responses to essay number 3, the retention essay.

Participant essay:

I would not participate in the strike because I am not a fan of unionized plants and I would be happy to have work at the time because there were several people¹ who would have loved to be working at the time. I would be very pleased with the 40cents an hour² because those were fairly competitive⁻¹ wages at the time because this took place during the depression.

Participant essay:

I would join the strike, because I do not believe that these workers were treated fairly at all. They were pushed to the limits of work and never given any breaks¹ for it.

Appendix F: Personal Data

Gender:

Female	
Male	

Use of information:

I have used a library catalog such as MAGIC before to find books, magazines or journals.

Yes	
No	

How many books, journals, etc. have you checked out from a *library* in the most recently completed *semester*?

Please select one:

0-3	
4-6	
7 or more	

How many *hours per week* do you spend searching for information (rather than reading email) on the *Internet* including Google.com, Amazon.com, Magic (MSU Library catalog), etc?

Please select one:

0-3 hours	
3-6 hours	
6-10 hours	
10-15 hours	
15 or more hours	

Appendix G: Recruitment Form

PARTICIPANTS NEEDED FOR A RESEARCH PROJECT

Research Purpose: To evaluate the effectiveness of a website that contains digitized full text, images and audio materials on a 20th century event in U.S. History.

This is a **two part study** on hypermedia. In the **first** part, you will be asked to meet in either the Berkey or South Kedzie (Room 222 South) computer lab in order to examine a number of informational objects that include references to traditional books, digitized full text, audio, and images on an event in 20th century US history. You will be asked questions about the event. Then you will be asked to **examine a hypermedia system and listen to 20 audio recordings and read 4 text files** and tested on the content you studied. You will briefly be asked what you thought about the system. Next, you will be asked to complete a vocabulary test and lastly you will be asked questions about yourself such as your gender and your use of library resources. In the **second** part, you will be asked to complete an online survey (from any location you wish) one week after you complete part one. You will be asked some questions about what you recall concerning the 20th century U.S. history event. Your answers to these and all questions will be kept confidential.

- You will be asked questions about the content of what you see and hear.
- Please bring earphones if you have them otherwise earphones will be provided for you.
- All responses will be and will remain confidential.

Required:

- Self declared history majors attending Michigan State University who are at least 18 years of age.

You are not being evaluated, but the website is.

Dates and times are for April 4th – April 20th :

Wednesday, 1-3 or 4-6, or Friday, 10-noon, 1-3, or 3:30-5:30.

Interested participants should call (517) 214.2626 or email Amy at wellsat@msu.edu to arrange an appointment. **You will be paid \$25.00 for 2 to 2.5 hours** of your time and your confidentiality will be assured.

This research is being conducted by an investigator from the College of Education.

Appendix H: Informed Consent and Explanation Form

Hypermedia and learning

This is a **two part study** on hypermedia. In the first part, you will be asked to examine a number of informational objects that include references to traditional books, digitized full text, audio, and images on an event in 20th century US history. You will be asked questions about your knowledge of the event. Then you will be asked to **examine a hypermedia system and listen to 20 audio recordings and read 4 text files**. Then you will be tested on the content you studied. You will briefly be asked what you thought about the system. Next you will be asked to complete a vocabulary test and lastly you will be asked questions about yourself such as your gender and your use of library resources. In the second part, you will be briefly asked some questions about what you recall concerning the event. Your answers to these and all questions will be kept confidential.

The purpose of this study is to better understand how undergraduates learn from different displays of electronic information. Such an increased understanding can guide the development of future systems, which might enable students to retain facts or mentally integrate information more quickly and/or efficiently.

Your participation in this study will take approximately 2 to 2.5 hours. There are no known risks associated with participation in this study. Your identity will be entered into a database as unique ID numbers by the project investigator, Amy

Tracy Wells. Thereafter only unique IDs will be associated with individual responses. Names will not appear in the reports nor will they be associated with survey analysis or results. Your identity will be kept confidential to the maximum extent provided by law and will only be known to the project investigator.

For the first part of the study, which will take approximately 2 hours, **you will be compensated \$20.00**. Payments will be in the form of cash following your participation. For the second part of the study, which will take less than 10 minutes, **you will be compensated \$5.00**.

You should be aware that other than the \$25.00 payment, you may not personally or directly benefit from any of the procedures administered or from observed results (although you may benefit from learning some interesting Michigan history.)

Your participation in this study is voluntary. You may choose not to participate at all, refuse to answer certain questions, or discontinue your participation at any time without penalty. If you have *any* questions or concerns regarding your participation in this study, you may contact the Principal Investigator Raven McCrory at (517) 353-8565, by email at mccrory@msu.edu, or regular mail: 513G Erickson Hall, East Lansing, MI 48824 or Amy Tracy Wells by phone at (517) 214.2626, by email at wellsat@msu.edu, or regular mail: 71 University Drive, East Lansing, MI 48823.

I am at least 18 years of age and voluntarily agree to participate in this study.

Signed: _____ Date: _____

Appendix I: General Instructions

(These instructions were be read aloud and distributed to all participants.)

The purpose of this research is to better understand how undergraduate history majors' learn here at MSU and elsewhere. The material you are about to read and hear about concerns the Flint Sit-Down Strike.

Your best effort is important and will benefit future students.

This is a two part study on hypermedia. Today you will complete part one. In part two, you will be asked to complete an online survey (from any location you wish) one week after you complete part one, which will take approximately 10 minutes. I will now hand you a number, 1-32. You will use this number today and for part two. Next week I will send you an email with the location of the second, brief survey. After you complete part two, send me email with your address and I will mail you \$5.00 cash or check - your pick.

Your answers to these and all questions will be kept confidential.

Here are the steps you will go through in this study. First you will complete a quick multiple choice test on the Flint Sit-Down Strike. Next and most important you will be exposed to a simple computer system that will allow you to access information about the Flint Sit-Down Strike. I will tell you more about how the system works in a few minutes. Before doing that I would like to give you the

overall picture of what you will be doing. You listen to a total of 20 audio files and read a total of 4 text files about the Flint strike. Afterwards you will be asked some factual questions and to write two essays based on the content you have been studying so be sure to listen and read very carefully. There will be a lot of information - try to put it together as best you can. Lastly you will complete a short survey about your experience, complete a two-part vocabulary test and answer a few questions about yourself.

Because much of the content is audio, you'll need to wear the headphones at all times.

Appendix J: Specific Instructions

(System specific instructions will be read aloud and distributed to all participants.)

Traditional library catalog

Please work independently of everyone around you but do raise your hand if you have a question. Thank you for your help and your best effort on the tasks ahead!

This system is similar to one you might find in a library. However instead of containing information about many subjects, it contains information only on the Flint Sit-Down strike. Remember your goal as you explore this system is to learn as much as you can about the Flint strike in preparation for a short series of tests you will complete.

All of the content is related to one of four themes: Working conditions, Strike methodology, Conditions during the Strike and Community Response. You have been given a list (Appendix Ic: Resources) of all the items in the systems that are grouped according to theme. (Note: some content is included in two different themes.) **You must listen to a total of 20 audio resources** (e.g., Leo Connelly interview concerning "Wages and Breaks at General Motors") **and read a total of four text resources** (e.g., Samuel Romer's synopsis of "wages and working conditions").

Just as with a traditional library catalog you can search by title, keyword, type of material, subject heading and/or author. After making a choice, you need to type in a word (e.g., "wages") or a phrase (e.g., "Working Conditions Caused the Strike"). Next you will see a list of items that contain or are related to the word(s) you used. Clicking on an item from the list brings up a screen that provides more information about the resource. Click on "connect to" to hear, see and/or read about the resources themselves.

You can always change your method of searching by choosing a different method from the bottom of the display or by clicking "new search". In addition, you can use "go forward" and "go backward" browser buttons to navigate. If you have a question, raise your hand. Thank you for your help and best effort on the tasks ahead. Remember to listen to a total of 20 audio resources and to read a total of 4 text resources – no more and no less. Please begin.

Simultaneous, non-linear system

Please work independently of everyone around you but do raise your hand if you have a question. Thank you for your help and your best effort on the tasks ahead!

The system is similar to many websites. However instead of containing information about many subjects, it contains information only on the Flint Sit-Down strike. Remember your goal as you explore this system is to learn as much as you can about the Flint strike in preparation for a short series of tests you will complete.

All of the content is related to one of four themes: Working conditions, Strike methodology, Conditions during the Strike and Community Response. You have been given a list (Appendix Ic: Resources) of all the items in the systems that are grouped according to theme. (Note: some content is included in two different themes.) **You must listen to a total of 20 audio resources** (e.g., Leo Connelly interview concerning "Wages and Breaks at General Motors") **and read a total of four text resources** (e.g., Samuel Romer's synopsis of "wages and working conditions").

You will have a lot of choices on how to proceed. To begin, you need to decide whether you will decide which topics you will browse: Working Conditions Before the Strike, Strike Methodology, Conditions During the Strike or Community

Response to the Strike. Next you will see a list of items that contain or are related to the topic you chose. Clicking on an item from the list brings up a screen that provides access to that resource as well as some information about the resource.

You can always browse another topic about something else from the bottom of the display or by clicking "new search". In addition, you can use "go forward" and "go backward" browser buttons to navigate. If you have a question, raise your hand. Thank you for your help and best effort on the tasks ahead. Remember to listen to a total of 20 audio resources and to read a total of 4 text resources – no more and no less. Please begin.

Appendix K: Resources

You must listen to a total of 20 audio resources (e.g., Leo Connelly interview concerning "Wages and Breaks at General Motors") and read a total of four text resources (e.g., Samuel Romer's synopsis of "wages and working conditions").

Note how many of the resources you listen to and read below (e.g., Text files read: IIII)

Audio files listened to:

Text files read:

	Author	Title (Primary or Secondary)	Type of Object
Theme: Working Conditions			
2	Connelly, Leo	Wages and Breaks at General Motors	Sound
3	Erlich, Ed	Piece Work at General Motors	Sound
4	Gage, Russell	Assembly Line Work: No Time for Water	Sound
5	Gancsos, Louis	Working Conditions Caused the Strike	Sound
6	Havrilla, Andrew	Wage System	Sound
7	Holland, Ray	GM and Turnover Rates	Sound
8	Jones, Larry	Working in the Heat	Sound
9	Jordan, Francis	Working the Line	Sound
10	K., Gillian	No Eating on the Job	Sound
11	Knotts, Ray	Demands of the Union	Sound
12	Linder, Walter	Working Conditions	Text
13	Lischer, Clarence	The Company Union	Sound
14	Mundale, Maynard	Getting Sick on the Line	Sound
15	Ricks, Grant	Pay Differences for the Same Work	Sound
16	Robinson, Leo	Lost Wages	Sound
17	Romer, Samuel	Wages and Working Conditions	Text
18	Schmitz, Peter	General Public Caught by Surprise	Sound
19	Skunda, Joseph	Assembly Line Work: No Time for Water Part II	Sound
20	The Detroit News	Flint Employees	Image

Appendix K (cont'd).

21	The Detroit News	Men at Work - Flint Interior	Image
22	The Detroit News	Men at Work - Flint Interior (2)	Image
23	The Detroit News	Men at Work - Flint Interior (3)	Image
24	The Detroit News	Men at Work - Flint Interior (4)	Image
Theme: Strike Methodology			
1	Adamic, Louis	Sitdown	Text
2	Gancsos, Louis	Strike Leaders Were More Radical	Sound
3	Gillette, Evelyn	Sit Down or Walk Out?	Sound
4	Jones, Larry	Secrecy of the Timing	Sound
5	Jones, Larry	The Role of Women at Chevrolet, Fisher Body and in the Red Berets	Sound
6	Kraus, Henry	The Theory of the Sit-Down Strike	Sound
7	Linder, Walter	Sit-Down Methodology	Text
8	Reider, Alexander	Keeping the Company out of the Plant During the Strike	Sound
9	Robin, Leo	John L Lewis and Frank Murphy Compromise	Sound
10	Root, Floyd	Effectiveness of the Sit-Down Strike	Sound
11	The Detroit News	National Guard Strike Duty (2)	Image
12	The Detroit News	In Front of Plant #1	Image
13	The Detroit News	Chevrolet Motor Car Company	Image
14	The Detroit News	Genora Johnson with a very young picketer	Image
15	The Detroit News	Bringing Food Under the National Guard	Image
16	Van Dyke, Michael	Strike Methodology	Text
Theme: Conditions During the Strike			
1	Erich, Ed	Sending and Receiving Messages	Sound
2	Fry, Joe	Food Sources During the Strike	Sound
3	Gillette, Evelyn	Sit Down or Walk Out?	Sound
4	Hubbard, Earl	Life During the Strike	Sound
5	Hubbard, Earl	Burning Burlap Instead of Coal	Sound
6	Jones, Larry	Secrecy of the Timing	Sound
7	Lovett, Robert Morss	A Stockholder Visits Flint	Text
8	Mcne, Sheldon	First Night of the Sit-in	Sound
9	Olay, Andrew	Shutting the Plant Down	Sound

Appendix K (cont'd).

10	Reider, Alexander	Keeping the Company out of the Plant During the Strike	Sound
11	Walker, Charles R.	Flint Faces Civil War	Text
12	The Detroit News	Sitting Down - Flint Interior	Image
13	The Detroit News	Man and Dog - Plant #1	Image
14	The Detroit News	Shaving - Flint, Interior View	Image
15	The Detroit News	Meal Time - Plant #1	Image
16	The Detroit News	Reading - Fisher Body	Image
17	The Detroit News	National Guard Strike Duty (2)	Image
Theme: Community Response			
1	Gadola, Mrs.	Police Protection for the Judge's Family	Sound
2	Gibbs, Robert	Local Churches Didn't Help the Strikers	Sound
3	Hayward, Laura	Flint Journal Labels Strikers as Reds or Socialists	Sound
4	Healy, Gerald	Police Consider and Reject Beating Up Strike Leaders	Sound
5	Healy, Gerald	Local Businesses Suffer During Strike	Sound
6	Kraus, Henry and Kraus Dorothy	Were Suspicious of Unions and Lewis	Sound
8	Loisell, Paul	Company Union was Sufficient	Sound
9	Lovett, Robert Morss	A Stockholder Visits Flint	Text
10	Moon, Mrs. Rollin	Role of the Company "Goons"	Sound
11	Ricks, Grant	Pay Differences for the Same Work	Sound
12	Schmitz, Peter	General Public Caught by Surprise	Sound
13	Van Dyke, Michael	Community Response	Text
14	The Detroit News	Strikers Overturning Car	Image
15	The Detroit News	Demonstration at No. 2 Plant	Image
16	The Detroit News	National Guard strike duty at Flint	Image
17	The Detroit News	National Guard Strike Duty (2)	Image
18	The Detroit News	Throwing Gas	Image

Appendix L: Debriefing Form

Thank you for your participation in the study! The study was designed to determine the effects of two different hypermedia displays on people's ability to recall and synthesis information.

This type of research is helpful as educators and system designers struggle with providing information in a manner that aids learners but also meets budgetary needs. This research was undertaken as part of the requirement to receive a doctorate in Education.

Appendix M: Post Test - Content knowledge of the Flint Sit-Down Strike

1. Prior to the Flint sit-down strike, General Motors had:
Welcomed the UAW in its plants
Attempted to halt union membership drives in its plants
Engaged with the UAW in collective bargaining
2. The decision to actually begin the strike was made by a majority of unionized workers.
True
False
3. The most common complaint that striking Flint workers voiced was that GM:
Refused to provide medical insurance
Did not provide a pension plan
Pushed constantly to speed-up production
3. The local newspaper, the Flint Journal, supported the Strike by making the larger Flint area aware of the workers' difficulties.
True
False
4. The sit-down strike was necessary in order to:
Prevent GM from removing the auto body dies
Avoid GM from hiring replacement workers
Prevent the violence normally associated with conventional strikes
All of the above
5. The National Guard was called up during the strike to prevent both parties from harming one another.
True
False
6. When was the Flint Sit-Down Strike?
1937-1938
1936-1937
1938-1939
7. Key support for the strikers was provided not just by families of the strikers but also by the local churches.
True
False
8. Which five of the following people were involved in the Flint Sit-Down Strike?
 - John L. Lewis
 - Bob Travis
 - Genora Johnson
 - Frank Murphy
 - Walter Reuther

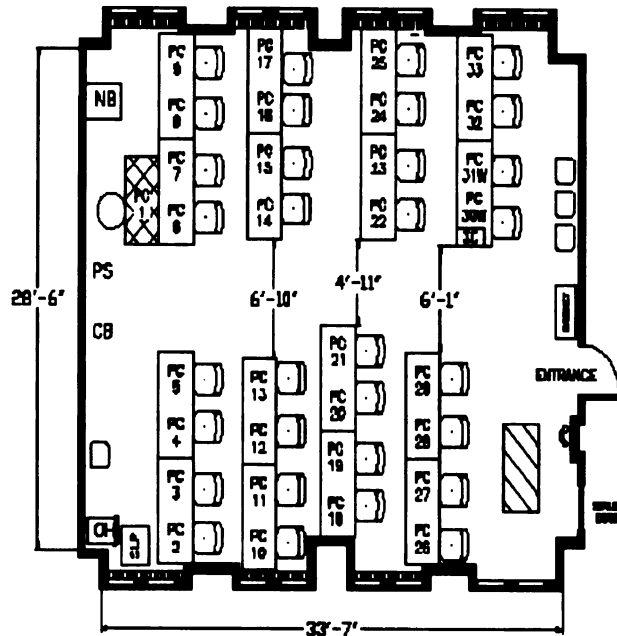
- Fania Reuther
- Stephen Bechtel
- David Samoff
- Thomas Watson Jr
- Bernard L. Beaubien
- G.W. Singer
- James McNelly
- Arthur "Bud" Fletcher
- James P. Hoffa
- Jerry Lewis

9. Imagine it is the winter of 1936-37 and you are an employee in one of GM's Flint plants. A fellow worker approaches you and asks you to participate in the sit-down strike that has just broken out in Fisher One and Two. Would you join in the strike? Using specific examples, name three reasons why you would or why you would not.

Appendix N: South Kedzie 222



222 S. KEDZIE COMPUTER LAB



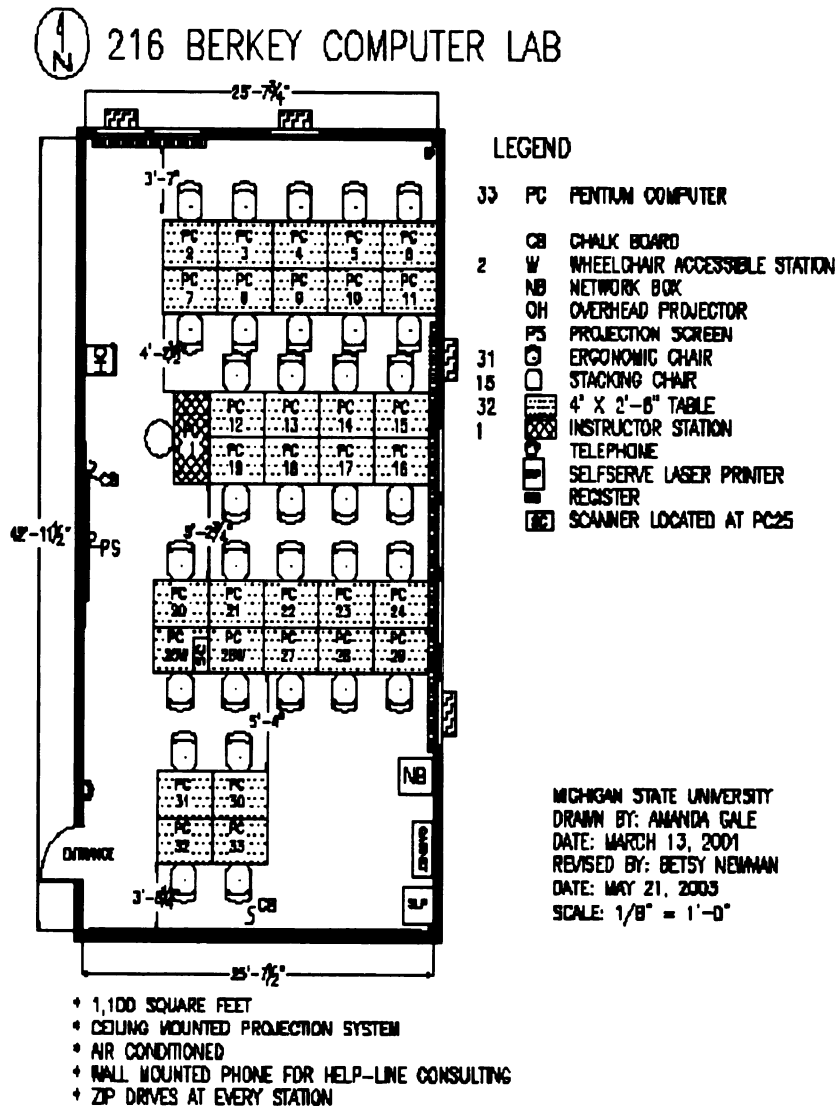
LEGEND

- | | | |
|----|----|-------------------------------|
| 33 | PC | PENTIUM COMPUTER |
| 2 | V | WHEELCHAIR ACCESSIBLE STATION |
| | NB | NETWORK BOX |
| | CB | CHALK BOARD |
| | OH | OVERHEAD PROJECTOR |
| | PS | PROJECTION SCREEN |
| | SS | SOFTWARE SERVER |
| 33 | | ERGONOMIC CHAIR |
| 18 | | STACKING CHAIR |
| 1 | | 5' X 2'-6" TABLE |
| 16 | | 6' X 2'-6" TABLE |
| | IS | INSTRUCTOR STATION |
| | | INSTRUCTOR STOOL |
| | | TELEPHONE |
| | | REGISTER |
| | | SELF SERVE LASER PRINTER |
| | | SCANNER LOCATED AT PC30 |

MICHIGAN STATE UNIVERSITY
 DRAWN BY: ERIKA PUTNAM
 DATE: NOVEMBER 19, 1998
 REVISED BY: ASHLEY BOWDISH
 DATE: APRIL 06, 2007
 SCALE: 1/8" = 1'-0"

- 1,106 SQUARE FEET
- CEILING MOUNTED PROJECTION SYSTEM
- AIR CONDITIONED
- WALL MOUNTED PHONE FOR HELP-LINE CONSULTING
- ZIP DRIVES AT EVERY STATION

Appendix O: Berkey 216



Appendix P: Specifics of Each Participant

This dissertation collected quantitative information on participant abilities and behaviors. Specifics for each person follow. Individuals who were in the control condition have asterisks next to their participant number.

	Number of participant									
Participant number	1*	2*	3*	4*	5*	6*	7*	8*	9*	10*
Testing Date										
(2007)	4/4	4/4	4/4	4/4	4/4	4/4	4/6	4/6	4/6	4/6
Total T/F & MC Correct	2	5	3	2	5	6	4	3	5	3
Total Labor Leaders Correct	2	2	1	3	2	3	1	2	2	3
Total Background Score	2.4	5.4	3.2	2.6	5.4	6.6	4.2	3.4	5.4	3.6

Appendix P (cont'd).

Total Recall

Score	3	4	0	5	5	4	7	6	1	12
--------------	----------	----------	----------	----------	----------	----------	----------	----------	----------	-----------

Total Integration

Score	12	9	7	6	11	8	9	10	8	8
--------------	-----------	----------	----------	----------	-----------	----------	----------	-----------	----------	----------

Word Count

Integration

Essays	303	256	269	266	253	267	293	259	312	295
---------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------

Total Synthesis

Score	5	3	2	3	5	4	3	5	4	1
--------------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------

Word Count

Synthesis Essay	291	256	255	285	281	263	297	408	208	250
------------------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------

Total Attitude	3.86	4.86	4.71	5	4.29	5.43	5.43	5.14	4.71	5
-----------------------	-------------	-------------	-------------	----------	-------------	-------------	-------------	-------------	-------------	----------

Total Vocabulary

Correct	20	24	16	16	26	20	19	27	17	22
----------------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------

Appendix P (cont'd).

What is

your

gender?

(Male or	Mal	Femal	Mal	Femal	Femal	Mal	Mal	Mal	Mal	Femal
Female)	e	e	e	e	e	e	e	e	e	e

I have

used a

library

catalog

such as

MAGIC

before to

find

books,

magazin

es or

journals.

(Yes or

No)	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
-----	-----	-----	-----	----	-----	-----	-----	-----	-----	-----

Appendix P (cont'd).

How many books, journals, etc. have you
checked out from a library in the most recently
completed semester? (0 to 3 (1), 3 to 6 (2), 7 or
more (3))

2 3 2 1 3 1 1 3 3 1

Appendix P (cont'd).

How many hours per week do you spend
searching for information (rather than
reading email) on the Internet including
Google.com, Amazon.com, Magic (MSU
Library catalog), etc? (0-3 hours (1), 3-6
hours (2), 6-10 hours (3), 10-15 hours

(4), 15 or more hours (5))	2	2	3	1	5	2	1	1	3	1
Total Background Correct	5	7	5	6	6	7	7	7	5	5
Total LL Correct	2	3	3	2	3	3	3	4	3	3
Total Retention Score	4	4	1	1	4	4	2	4	4	4

Appendix P (cont'd).

	Number of participant									
Number of participant	11	12	13	14	15	16	17	18	19	20
Testing Date										
(2007)	4/6	4/6	4/6	4/11	4/11	4/11	4/11	4/11	4/11	4/11
Total T/F & MC										
Correct	5	4	4	3	4	6	3	5	4	5
Total Labor										
Leaders Correct	2	1	1	1	3	0	3	2	2	1
Total Background										
Score	5.4	4.2	4.2	3.2	4.6	6	3.6	5.4	4.4	5.2
Total Recall Score	9	1	0	2	5	1	2	4	4	3

Appendix P (cont'd).

Total						Not			Not	
Integration						scored			scored	
Score	10	9	7	9	11	^	8	9	^	11
Word										
Count										
Integration					30					
Essays	250	267	349	264	2	51	272	368	52	220
Total						Not			Not	
Synthesis						scored			scored	
Score	6	3	4	4	6	^	2	5	^	4
Word										
Count										
Synthesis					36					
Essay	288	239	339	252	4	55	283	382	62	285
Total	4.1	4.1	4.8	4.8			4.5	3.8		5.1
Attitude	4	4	6	6	5	5.29	7	6	5.14	4
Total										
Vocabulary										
y Correct	25	12	24	25	26	15	18	31	15	23

Appendix P (cont'd).

What is

your

gender?

(Male or	Fema	Mal	Fema	Fema	Fema	Fema	Mal	Mal	Mal	Mal
----------	------	-----	------	------	------	------	-----	-----	-----	-----

Female)	le	e	le	le	le	le	e	e	e	e
---------	----	---	----	----	----	----	---	---	---	---

I have

used a

library

catalog

such as

MAGIC

before to

find

books,

magazin

es or

journals.

(Yes or

No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Appendix P (cont'd).

How many books, journals, etc.

have you checked out from a library

in the most recently completed

semester? (0 to 3 (1), 3 to 6 (2), 7 or

more (3))

3 1 1 1 3 3 3 3 1 3

How many hours per week do you

spend searching for information

(rather than reading email) on the

Internet including Google.com,

Amazon.com, Magic (MSU Library

catalog), etc? (0-3 hours (1), 3-6

hours (2), 6-10 hours (3), 10-15

hours (4), 15 or more hours (5))

4 1 3 3 2 3 1 3 5 2

No

No

Total Background Correct

score# 5 7 5 7 score# 7 6 5 7

Total LL Correct

No

No

score# 3 2 4 3 score# 3 3 2 3

No

No

Total Retention Score

score# 4 3 3 4 score# 1 1 0 4

Appendix P (cont'd).

Number of participant										
Number of particip ant	21*	22*	23	24	25	26	27	28*	29*	30
Testing	4/13/	4/13/	4/13/	4/13/	4/13/	4/13/	4/13/	4/20/	4/20/	4/20/
Date	07	07	07	07	07	07	07	07	07	07
Total										
T/F & MC										
Correct	3	5	4	2	6	6	1	2	3	3
Total										
Labor										
Leaders										
Correct	1	3	2	3	2	1	0	2	3	2
Total										
Backgr ound										
Score	3.2	5.6	4.4	2.6	6.4	6.2	1	2.4	3.6	3.4
Total										
Recall										
Score	7	10	5	3	0	1	0	2	3	1

Appendix P (cont'd).

Total Integration

Score	8	11	9	11	9	9	7	8	7	5
--------------	----------	-----------	----------	-----------	----------	----------	----------	----------	----------	----------

Word Count

Integration

Essays	286	268	251	278	272	262	312	297	266	408
---------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------

Total Synthesis

Score	6	4	2	4	4	2	5	5	3	3
--------------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------

Word Count

Synthesis Essay	421	268	305	266	292	301	333	441	339	327
------------------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------

Total Attitude	5.57	5.57	5.86	4.29	5.14	5.43	5.43	5.29	5	6
-----------------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	----------	----------

Total Vocabulary

Correct	16	30	32	21	20	26	17	14	17	14
----------------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------

Appendix P (cont'd).

What is

your

gender?

(Male or	Fem	Fem	Mal	Fem	Mal	Mal	Fem	Fem	Fem	Fem
----------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Female)	ale	ale	e	ale	e	e	ale	ale	ale	ale
---------	-----	-----	---	-----	---	---	-----	-----	-----	-----

I have

used a

library

catalog

such as

MAGIC

before

to find

books,

magazi

nes or

journals

. (Yes

Ye

Ye

Ye

or No)

Yes

Yes

s

Yes

s

s

Yes

Yes

Yes

Yes

Appendix P (cont'd).

How many books, journals, etc. have you
checked out from a library in the most recently
completed semester? (0 to 3 (1), 3 to 6 (2), 7
or more (3))

3 3 3 3 3 2 2 1 1 3

How many hours per week do you spend
searching for information (rather than reading
email) on the Internet including Google.com,
Amazon.com, Magic (MSU Library catalog),
etc? (0-3 hours (1), 3-6 hours (2), 6-10 hours
(3), 10-15 hours (4), 15 or more hours (5))

2 2 5 3 3 3 2 2 2 2

Total Background Correct

7 5 7 6 5 6 2 7 2 4

Total LL Correct

3 3 4 2 1 3 1 2 2 2

Total Retention Score

2 4 4 1 3 4 0 2 3 3

Appendix P (cont'd).

Number of participant	Number of participant		
	31*	32*	33*
Testing Date	4/27/07	4/27/07	4/27/07
Total T/F & MC Correct	1	5	4
Total Labor Leaders Correct	3	2	1
Total Background Score	1.6	5.4	4.2
Total Recall Score	3	4	1
Total Integration Score	11	8	8
Word Count Integration Essay	455	259	264
Total Synthesis Score	5	2	2
Word Count Synthesis Essays	484	299	265
Total Attitude	4	5	4.86
Total Vocabulary Correct	25	19	16
What is your gender? (Male or Female)	Male	Female	Male
I have used a library catalog such as MAGIC before to find books, magazines or journals. (Yes or No)	Yes	Yes	Yes
How many books, journals, etc. have you checked out from a library in the most recently completed semester? (0 to 3 (1), 3 to 6 (2), 7 or more (3))	3	3	3

Appendix P (cont'd).

How many hours per week do you spend searching for information (rather than reading email) on the Internet including Google.com, Amazon.com, Magic (MSU Library catalog), etc? (0-3 hours (1), 3-6 hours (2), 6-10 hours (3), 10-15 hours (4), 15 or more hours (5))	4	2	1
Total Background Correct	6	7	4
Total LL Correct	4	3	1
Total Retention Score	3	4	3

^Two participants were removed from this analysis because their essays were less than seventy words. However, one participant's essay contained 240 words but was retained.

Two participants though they responded to the T/F and multiple choice questions did not provide responses to the short answer question meaning that a total of 29 short answers were evaluated.

REFERENCES

- Alexandria Digital Library (n.d.) Alexandria Digital Library. Retrieved March 13, 2007, from <http://webclient.alexandria.ucsb.edu/mw/index.jsp>
- Bates, M. (2003) Task force recommendation 2.3 Research and design review: Improving user access to Library catalog and portal information. Final Report. Library of Congress, June 2003, 58p.
<http://www.loc.gov/catdir/bibcontrol/2.3BatesReport6-03.doc.pdf>
- Bates, M.J. (2002) Toward an integrated model of information seeking and searching. *The New Review of Information Behavior Research*, 3, 1-15.
- Bates, M. J. (1989). The design of browsing and berrypicking techniques for online search interface. *Online Review*, 13, 407-424.
- Bera, S. & Liu, M. (2006). Cognitive tools, individual differences, and group processing as mediating factors in a hypermedia environment. *Computers in Human Behavior*, 22, 295-319.
- Bernard, M. L., Hull, S., & Chaparro, B. (2005). Examining the performance and preference of embedded and framed/non-framed hyperlinks. *International Journal of Industrial Ergonomics*, 35(2), 139-147.
- Bernard R.M., Abrami, P. C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., Wallet, P. A., Fiset, M., & Huang, B. (2004). How does distance education compare to classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research*, 74 (3), 379-439.
- Borgman, C. L. (2004). The Interaction of Community and Individual Practices in the Design of a Digital Library. *International Symposium on Digital Libraries and Knowledge Communities in Networked Information Society*, Ibaraki, Japan., University of Tsukuba.
- Borgman, C. L. (2000). *From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World*. Cambridge, MA: The MIT Press.
- Borgman, C.L. (1999). What are digital libraries? Competing visions [Special Issue] *Information Processing & Management*, 38(3), 227-243.
- Borgman CL (1996). Why are online catalogs still hard to use? *Journal of the American Society for Information Science and Technology*. 47(7), 493-503.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How People Learn: Brain, Mind, Experience, and School* (Expanded edition). Washington D.C.: National Academy Press.

- Brinkerhoff, J. D., Klein, J. D., & Koroghlanian, C. M. (2001). Effects of overviews and computer experience on learning from hypertext. *Journal of Educational Computing Research*, 25(4), 427-440.
- Brunye', T.T., Taylor, H.A., Rapp, D. N. & Spiro, A.B. (2006). Learning Procedures: The Role of Working Memory in Multimedia Learning Experiences. *Applied cognitive psychology*, 20(7), 917-940.
- Calandra, B., & Barron, A. E. (2005). A preliminary investigation of advance organizers for a complex educational website. *Journal of Educational Multimedia and Hypermedia*, 14(1), 5 -23.
- Chen, S.Y., Ghinea, G., & Macredie, R.D.(2006). A cognitive approach to user perception of multimedia quality: An empirical investigation. *International Journal of Human-Computer Studies*, 64 (2006) 1200–1213
- Clark, J. M. & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3(3), 149-170.
- Cotter, James (2006). Class of 2006 Report. Michigan State University Retrieved on April 3, 2007, from http://admissions.msu.edu/documents/Class_of_2006_Report.pdf
- Dee-Lucas, D., & Larkin, J. H. (1999). Hypertext segmentation and goal compatibility: Effects on study strategies and learning. *Journal of Educational Multimedia and Hypermedia*, 8(3).
- Detroit News (n.d.) The historic 1936-37 Flint auto plant strikes. Retrieved May 3, 2006, from <http://info.detnews.com/history/story/index.cfm?id=115&category=business>
- Dillon, A., & Jobst, J. E. (2005). Multimedia Learning with Hypermedia. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 569-588). New York, NY: Cambridge University Press.
- Dillon, A., & Gabbard, R. (1998). Hypermedia as an Educational Technology: A Review of the Quantitative Research Literature on Learner Comprehension, Control, and Style. *Review of Educational Research*, 68(3), 322-349.
- Drabenstott, K.M. (1991). Online catalog user needs and behavior. Think Tank on the Present and Future of the Online Catalog: Proceedings. *RASD Occasional Papers*, No. 9. Chicago: American Library Association.

- Ellis, D. (1989). A behavioral approach to information retrieval system design. *Journal of Documentation*, 45(3), 171-212.
- Ekstrom, R.B. French, J.W. & Harman, H.H. (1976). Advanced Vocabulary Test I—V-4. *Manual for Kit of Factor Referenced Cognitive Tests*. Princeton, NJ: Educational Testing Service.
- Ericsson, K.A. and Simon, H.A. (1980). Verbal reports as data. *Psychological Review*, 87(3), 215-251.
- Eveland, W. P. J., Marton, K., & Seo, M. (2004). Moving Beyond "Just the Facts": The Influence of Online News on the Content and Structure of Public Affairs Knowledge. *Communication Research*, 31(1), 82-108.
- Ford, N., & Chen, S. Y. (2000). Individual Differences, Hypermedia Navigation, and Learning: An Empirical Study. *Journal of Educational Multimedia and Hypermedia*, 9(4), 281-311.
- Gauss, B., & Urbas, L. (2003). Individual Differences in Navigation between Sharable Content Objects—An Evaluation Study of a Learning Module Prototype. *British Journal of Educational Technology*, 34(4), 499-509.
- Hofman, R., & van Oostendorp, H. (1999). Cognitive Effects of a Structural Overview in a Hypertext. *British Journal of Educational Technology*, 30(2), 129-140.
- Howell, D. (2002). *Statistical methods for psychology* (5th ed.) Belmont, CA: Duxbury Press.
- Huk, T., & Steinke, M. (2007). Learning cell biology with close-up views or connecting lines: Evidence for the structure mapping effect. *Computers in Human Behavior*, 23(3), 1089-1104.
- Jacobson, M. & Spiro, R. (1995). Hypertext learning environments, cognitive flexibility, and the transfer of complex knowledge: An empirical investigation. *Journal Educational Computing Research*, 12(4) 301-333.
- Jonassen, D. H., & Wang, S. (1993). Acquiring structural knowledge from semantically structured hypertext. *Journal of Computer-Based Instruction*, 20(1), 1-8.
- Kozma, R. B. (1987). The implications of cognitive psychology for computer-based learning tools. *Educational Technology*, 28(11), 20-25.
- Lee, M.J. (2005). Expanding hypertext: Does it address disorientation? Depends on individuals' adventurousness. *Journal of Computer-Mediated*

Communication, 10(3), article 6. Retrieved on April 3, 2007, from <http://jcmc.indiana.edu/vol10/issue3/lee.html>

Lee, M. J., & Tedder, M. C. (2004). Introducing Expanding Hypertext Based on Working Memory Capacity and the Feeling of Disorientation: Tailored Communication Through Effective Hypertext Design. *Journal of Educating Computer Research*, 30(3), 171-195.

Lee, M. J., & Tedder, M. C. (2003). The effects of three different computer texts on readers' recall: Based on working memory capacity. *Computers in Human Behavior*, 19(6), 767-783.

Liaw S-S, Huang H-M, Chen G-D. (2007) An activity-theoretical approach to investigate learners' factors toward e-learning systems. *Computers in Human Behavior*, 23,1906-1920.

Library of Congress (n.d.). The Library of Congress: American Memory. Retrieved March 13, 2006, from <http://memory.loc.gov/ammem/index.html>

Liu, M. (2006). The Effect of a Hypermedia Learning Environment on Middle School Students' Motivation, Attitude, and Science Knowledge, *Computers in the Schools*, 22(3-4), 159-171.

Liu, M., Bera, S., Corliss, S. B., Svinicki, M. D., & Beth, A. D. (2004). Understanding the Connection between Cognitive Tool Use and Cognitive Processes as Used by Sixth Graders in a Problem-Based Hypermedia Learning Environment. *Journal of Educational Computing Research*, 31(3), 309 - 334.

Lubetsky, S. (1969). Principles of cataloging: Final Report. Phase 1: Descriptive cataloging. University of California, Los Angeles, Institute of Library Research.

Mayer, R. E. (2001). *Multimedia Learning*. Cambridge: Cambridge University Press.

Mitchell, T. J. F., Chen, S. Y., & Macredie, R. D. (2005). Hypermedia Learning and Prior Knowledge: Domain Expertise vs. System Expertise. *Journal of Computer Assisted Learning*, 21(1), 53-64.

Mishra, P., Spiro, R. J. & Feltovich, P. (1996). Technology, representation & cognition. In von Oostendorp, H. (Ed.) *Cognitive aspects of electronic text processing*. Norwood, NJ: Ablex Publishing Corporation.

- Nadolski, R. J., Kirschner, P. A., & van Merriënboer, J. J. G. (2006). Process Support in Learning Tasks for Acquiring Complex Cognitive Skills in the Domain of Law. *Learning and Instruction*, 16(3), 266-278.
- National Science Digital Library (n.d.) National Science Digital Library. Retrieved March 13, 2007, from <http://nsdl.org/>
- Nielsen, J. (2000). Why You Only Need to Test With 5 Users. Retrieved July 19, 2007, <http://www.useit.com/alertbox/20000319.html>.
- Nielsen, J. (1994). Heuristic Evaluation. In *Usability inspection methods*, J. Nielsen and R.L. Mack. New York: John Wiley & Sons.
- Nielsen, J. (n.d.). How to Conduct a Heuristic Evaluation. Retrieved July 19, 2007, from http://www.useit.com/papers/heuristic/heuristic_evaluation.html
- Nielsen, J. (n.d.) *How to Conduct a Heuristic Evaluation*. Retrieved July 19, 2007, http://www.useit.com/papers/heuristic/heuristic_evaluation.html
- Pierotti, D. (1995) Heuristic Evaluation - A System Checklist. Retrieved July 19, 2007, from <http://www.stcsig.org/usability/topics/articles/he-checklist.html>
- Oostendorp, H. V., & Nimwegen, C. V. (1998). Locating information in an online newspaper. *Journal of Computer Mediated Communication*, 4(1), Np.
- Pavio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.
- Pettigrew, K. E., Fidel, R. & Bruce, H. (2001) Conceptual Frameworks in Information Behavior. *Annual review of information science and technology* (ARIST), 35, p. 43-78.
- Puntambekar, S., & Stylianou, A. (2005). Designing navigation support in hypertext systems based on navigation patterns. *Instructional Science*, 33(5-6), 451-481.
- Rapp, D.N. Taylor, H.A. and Crane, G.R. (2003). The impact of digital libraries on cognitive processes: Psychological issues of hypermedia. *Computers in Human Behavior*, 19 (5), 609-628.
- Rice, R. E., McCreddie, M. & Chang, S-J L. (2001). The Importance of Accessing and Browsing. *Accessing and browsing information and communication*. Cambridge, MA: The MIT Press.

- Sandberg-Fox, A.M. (Ed.) (2001). *Proceedings of the Library of Congress Bicentennial Conference on Bibliographic Control for the New Millennium*. Washington, D.C.: Library of Congress Cataloging Distribution Service.
- Song, S.H. & Keller, J.M. (2001) Effectiveness of Motivationally Adaptive Computer-Assisted Instruction on the Dynamic Aspects of Motivation. *ETR&D*, 49(2), 5–22.
- Spiro, R.J.; Collins, B.P.; Thota, J.J.; & Feltovich, P.J. (2004). Cognitive Flexibility Theory: Hypermedia for Complex Learning. *Educational Technology*, 43(5), 5-12.
- Spiro, R.J., Feltovich, P.J., & Coulson, R.L. (1996). Two epistemic world-views: Prefigurative schemas and learning in complex domains. *Applied-Cognitive-Psychology*, 10, S51-S61.
- Spiro, R., Feltovich, P.J., Jacobson, M.J. & Coulson, R.L. (1992) "Knowledge representation, content specification, and the development of skill in situation-specific knowledge assembly: Some constructivist issues as they relate to cognitive flexibility theory and hypertext." In T. M. Duffy & D. H. Jonassen (Eds.) *Constructivism and the technology of instruction: A conversation* (pp. 57-76). Mahwah, NJ: Lawrence Erlbaum Assoc Inc.
- Spiro, R.J. & Jehng, J-C. (1990). Cognitive Flexibility and Hypertext: Theory and Technology for the Nonlinear and Multidimensional Traversal of Complex Subject Matter. Nix, D. & Spiro, R.J. (Eds). *Cognition, education, and multimedia: Exploring ideas in high technology*. (pp. 163-205). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.
- Spiro, R.J. Coulson, R.L., Feltovich, P.J., & Anderson, D.K. (1988). Cognitive flexibility theory: advanced knowledge acquisition in ill-structured domains. *Tenth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Assoc Inc.
- Spiro, R.J., Vispoel, WP, Schmitz, JG, Samarapungavan, A, & Boerger, AE. (1987). Knowledge Acquisition for Application - Cognitive Flexibility and Transfer in Complex Content Domains. Britton, B.K. & Glynn, S.M. (Eds). (1987). *Executive control processes in reading. Psychology of reading and reading instruction*. (pp. 177-199). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.
- Su, Y., & Klein, J. D. (2006). Effects of navigation tools and computer confidence on performance and attitudes in a hypermedia learning environment. *Journal of Educational Multimedia and Hypermedia*, 15(1), 87-106.

- SurveyMonkey (n.d.). SurveyMonkey. Retrieved March 13, 2006, from <http://surveymonkey.com/>
- Schwartz, N.H., Andersen, C., Hong, N., Howard, B. & McGee, S. (2004). The influence of metacognitive skills on learners' memory of information in a hypermedia environment. *Journal of educational computing research*, 31(1) 77-93.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257-285
- Triantafillou, E., Pomportsis, A., Demetriadis, S., & Georgiadou, E. (2004). The value of adaptivity based on cognitive style: An empirical study. *British Journal of Educational Technology*, 35(1), 95-106.
- Tardieu, H., & Gyselinck, V. (2003). Working memory constraints in the integration and comprehension of information in a multimedia context. In H. van Oostendorp (Ed.), *Cognition in a digital world* (pp. 3-24). Mahwah, NJ: Lawrence Erlbaum Associates.
- Thomas, S. E. (2001). The Catalog as portal to the internet. *Proceedings of the Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web* (pp. 21-37). Washington, DC: Library of Congress.
- Waniek, J., Brunstein, A., Naumann, A., & Krems, J. F. (2003). Interaction between text structure representation and situation model in hypertext reading. *Swiss Journal of Psychology Schweizerische Zeitschrift fuer Psychologie Revue Suisse de Psychologie*, 62(2), 103-111.
- Wells, A.T. (in press). Learning objects: A graduate student perspective. *The digital puzzle*. Chicago, IL: The American Library Association.
- Wells, A.T. (2005). In Search of Learning and Knowing. Posted session presented at the annual meeting of the American society of information science and technology, Charlotte, NC.
- Yang, S.C. (2001). An interpretive and situated approach to an evaluation of Perseus digital libraries. *Journal of the American Society for Information Science and Technology*, 52, 1210-1223.
- Zahn, C., Barquero, B., & Schwan, S. (2004). Learning with hyperlinked videos--design criteria and efficient strategies for using audiovisual hypermedia. *Learning and Instruction*, 14(3), 275-291.

- Zumbach, J. (2006). Cognitive Overhead in Hypertext Learning Reexamined: Overcoming the Myths. *Journal of Educational Multimedia and Hypermedia*, 15(4), 411-432.
- Zumbach, J., Reimann, P., & Koch, S. (2001). Influence of passive versus active information access to hypertextual information resources on cognitive and emotional parameters. *Journal of Educational Computing Research*, 25(3), 301-318.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 02956 9864