The Effect of WWW Document Structure on Students' Information Retrieval

Ian Brown

Abstract:

This experiment investigated the effect the structure of a WWW document has on the amount of information retained by a reader. Three structures common on the Internet were tested: one long page; a table of contents leading to individual sections; and short sections of text on separate pages with revision questions. Participants read information structured in one of these ways and were then tested on recall of that information. A further experiment investigated the effect that 'browsing' – moving between pages – has on retrieval. There was no difference between the structures for overall amount of information retained. The single page version was best for recall of facts, while the short sections of text with revision questions led to the most accurate inferences from the material. Browsing on its own had no significant impact on information retrieval. Revision questions rather than structure per se were therefore the key factor.

Keywords:

Document structure, learning, memory traces, revision.

Commentaries:

All JIME articles are published with links to a commentaries area, which includes part of the article's original review debate. Readers are invited to make use of this resource, and to add their own commentaries. The authors, reviewers, and anyone else who has 'subscribed' to this article via the website will receive email copies of your postings.

Demonstrations:

The websites contrasted in the experiments are explained in Sections 1.1.1-1.1.3 (Expt. 1), and 3.2 (Expt. 2), with links to the respective websites.

Ian Brown, *Department of Computer Science, University College London, London WC1E 6BT, U.K.*

I.Brown@cs.ucl.ac.uk http://www.cs.ucl.ac.uk/staff/I.Brown/

1. Introduction

1.1 Rationale

Hypertext is a computing system first proposed by Vanuvar Bush (1945). It consists of a series of on-line texts linked by 'hotspots' - important words or phrases which are highlighted in the text. By clicking on these hotspots, users can 'jump' between sections of the text.

The World-Wide Web (WWW) has greatly popularised hypertext. It was designed as an information dissemination tool, but educational materials are migrating to the WWW at great speed. Good design for one purpose, however, may not be good for the other.

One of hypertext's most important design features is how information is structured. Does this affect how much students learn from hypertexts? At one extreme, the document may be little different from a book - one long section of text. At the other, individual paragraphs, with links between them, may take up a 'page' each. Sections may also include questions to encourage readers to review the material as it is read.

This experiment investigates whether this structure has an effect on the amount of information a reader can recall a short time later. Three structures common on the Internet were chosen as a representative range, as follows.

1.1.1 Unified text version

Paragraphs are the natural unit of text. Hypertext systems use them to create natural section breaks. Nash, however, cautioned that it is "a mistake to think of them as self-contained units, the 'building blocks' of text." (1980, p8). Arguments flow between paragraphs; splitting them up breaks the links which serve to reinforce the author's message. Peter Whalley (1993) argues strongly that serving up linear, cohesive texts as small chunks does not encourage deep learning. He cites Nash on the expository approach to writing:

There is a programme of assertions, examples, qualifications, but they are not presented as a series of distinctly labelled positions. Instead they are related to each other in a progressive unfolding pattern, the turns and connections of which are demonstrated in various ways. (p.11)

By splitting up a text into short, independent sections, a hypertext author loses this coherence and reinforcement. While hypertext is ideal for presenting small pieces of unrelated information, in an encyclopaedic form, educational materials are rarely structured in this manner. Whalley predicts that "the fragmentation effect in hypertext… is likely to make it more difficult for the learner to perceive the author's intended argument structure" (p.11).

This document structure therefore presents information as one unified page of text. The test version is linked to the web version of this document[1].

1.1.2 Hierarchical version

A hierarchical contents page leading to small sub-sections is the most common structure for WWW pages. Psychologists initially thought this would aid learning, by providing a ready-made semantic structure for learners to build information into (e.g. Fiderio, 1988) These structures represent learning schemas as hypertext pages, with ordered labelled relationships as hypertext links (Jonassen, 1993). In effect, they mimic the concept of a semantic network memory model.

Recent work has cast doubt on this view. Jonassen concluded that "merely providing structural clues in the user interface of a hypertext will not result in significant increases in structural knowledge acquisition." Dillon, McKnight and Richardson (1993) put it best:

Semantic space is an abstract psycholinguistic concept which cannot be directly observed... By definition, [it] is n-dimensional and practically unbounded... We must be clear that here we are not navigating through, or on the basis of, semantics. (p186)

Nevertheless, semantic memory is believed to be organised as a network model (Collins and Quillian, 1969). It is possible that this hypertext, which is ordered in a similar manner[2] , could improve the organisation of the material learned, and thus its recall.

This document structure is included to evaluate the usefulness of this extremely common hypertext design[3].

1.1.3 Active Review version

Craik and Lockhart's classic levels of processing theory postulates that 'deeper', more semantic processing of material leads to better learning (1972). Anderson and Reder (1979) proposed

---

[1]   *Unified version http://www-jime.open.ac.uk/98/12/demos/unified/*

[2]   *The hierarchical hypertext imitates almost exactly original hierarchical network model proposed by Collins and Quillian. Later work (Collins and Loftus, 1975) expands the concept to a more complex spreading activation model, which is similar to a hypertext with links spread throughout the document rather than simply as top-down connections. This could be another interesting avenue for research into hypertext and memory.*

[3]   *http://www-jime.open.ac.uk/98/12/demos/hierarchical/*

elaboration as one explanation of this. By processing material semantically, it is stored in many different, elaborated ways - improving subsequent recall. Walker, Jones and Marr (1983) suggested an additional mechanism: that increased cognitive effort (defined as the proportion of available cognitive resources consumed by a task) would increase recall, by creating a more distinctive memory trace.

This document structure encourages this elaboration and extra processing. After reading a short section of material, students are asked a question. This forces them to review and re-interpret the previous material - leading to more processing and more elaborate storage. If the question is answered correctly, the next section of material is shown. If it is not, an explanation of the correct answer is given before participants move on to the next page.

The test version is linked to the web version of this document[4].

1.2   Reproductive and reconstructive memory

Wickelgren (1972) makes an important distinction in learning. Reproductive memory stores simple facts. Reconstructive memory requires inferences to be made from learned material. While reproductive learning is factual, reconstructive learning is more semantic in nature. Samarapugavan and Beishuizen (1990), using a hypertext containing several different perspectives on the same material, found an improvement only in participants' reconstructive learning.

1.3   These experiments

Experiment 1, entitled *Comparing three common document structures* compares the three hypertext versions above. Experiment 2, entitled *The browsing effect* compares the unified and hierarchical versions to uncover any effect caused by browsing, where users have to flick backwards and forwards between small sub-sections of text presented individually. This is expected to disrupt concentration and Nash's (1980) expository flow, and therefore learning.

Reproductive and reconstructive questions were included in the tests participants took after studying one of the hypertexts, to uncover effects on both types of learning.

The hypothesis was that document structure *would* affect learning, in both overall scores and the separated reproductive and reconstructive scores. The active review hypertext was expected to be most successful, due to the increased elaboration and processing it causes. The hierarchical version was expected to be least successful, due to a detrimental browsing effect not counter-balanced by any semantic factors.

[4]   *http://www-jime.open.ac.uk/98/12/demos/active/*

2.   Experiment 1: Comparing three common document structures

2.1   Method

The 42 participants were new Computing Science students. While some had studied the subject, the majority had minimal computing experience. They were randomly selected during three 'computer introduction' sessions in a computing laboratory.

Participants were randomly assigned to one of three groups corresponding to the unified, hierarchical or active review hypertexts. The documents' contents concerned Joan Miro, a relatively obscure Spanish artist unknown to the students (this was checked with each participant). The hypertexts differed in structure and the presence or absence of revision questions. The pages were designed to take about two minutes to read, allowing students to review them in the allocated five minutes.

Netscape Navigator (a WWW browser) was set up on Macintosh LC475 computers at the relevant starting page.

Each participant was instructed to read the page(s) for five minutes with the intention of taking a paper-based test afterwards (in Appendix 1).[5] Students had already been taught how to use the browser.

After studying their assigned WWW document for five minutes, participants took a test containing seven reconstructive and thirteen reproductive randomly-ordered questions on the hypertext. Five minutes were allowed for the test.

Participants started the experiment at different times and sat at a distance from each other. The question sheet reminded subjects that the test results were anonymous, and asked for their co-operation in completing the questions without help. This was given in all cases. As a follow-up, a poster was displayed in the first-year computing area showing results and a WWW address for further information and queries.

[5]   *Numerous studies have shown intention to learn has little effect on learning (Hammond, 1993, p57). Students using Computer-Aided Learning (CAL) tools will (hopefully!) always intend to learn.*

2.2   Results

Analysis of participants' total test scores[6] gave the following results.

| Document structure	Mean score	Standard deviation
Unified (n=16)	6.071	3.872
Hierarchical (n=14)	6.500	.521
Active review (n=12)	5.833	1.467

Table 1: *Total test scores on the three document structures*

A one-way ANOVA test showed there was no significant difference between groups (F(2,41)=.16, p=0.856).

Scores were then separated into reproductive and reconstructive question totals.

2.2.1 Reproductive questions

Figure 1 and Table 2 show that the unified and hierarchical groups scored more (0.9857 and 1.002 standard deviations respectively) on the test than the active review group.
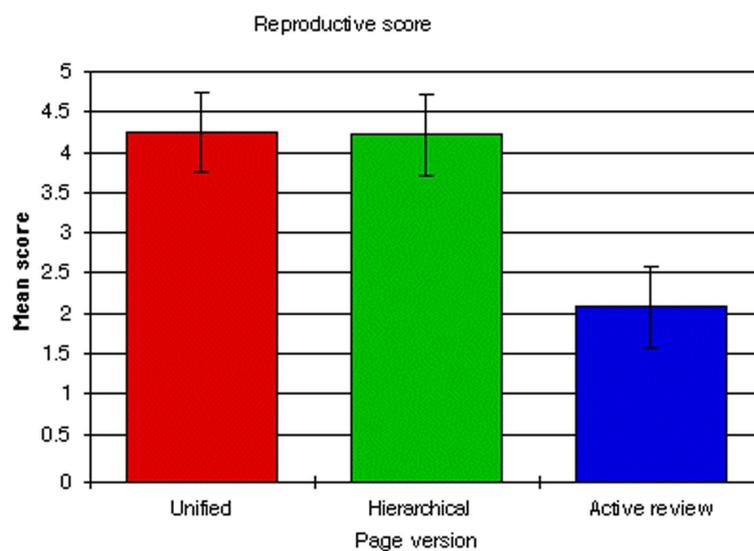
Figure 1: *Graph of scores on the reproductive questions for each web structure*

[6]   *Scores were corrected for guessing on questions 5, 6, 7, 12, 13 and 16 (where the answer was yes/no) by deducting one point for an incorrect answer.*

Document structure	Mean score	Standard deviation
Unified	4.250	2.569
Hierarchical	4.214	2.359
Active review	2.083	0.996

Table 2: *Mean scores and standard deviations on the reproductive questions for each web structure*

A one-way ANOVA test showed there was a significant difference between groups ($F_{(2,41)}=4.24$, $p=0.022$). A post hoc Scheffé test showed there was a significant difference in test score between the unified and active review groups ($F_{(2,41)}=3.25$, $p < .05$) and the hierarchical and active review groups ($F_{(2,41)}=3.26$, $p < .05$) but not between the unified and hierarchical groups ($F_{(2,41)}=-4.05$, $p > .05$).

2.2.2 Reconstructive questions

Figure 2 and Table 3 show that the active review group scored more on the test than the unified and hierarchical groups (0.8574 and 0.9889 standard deviations respectively).
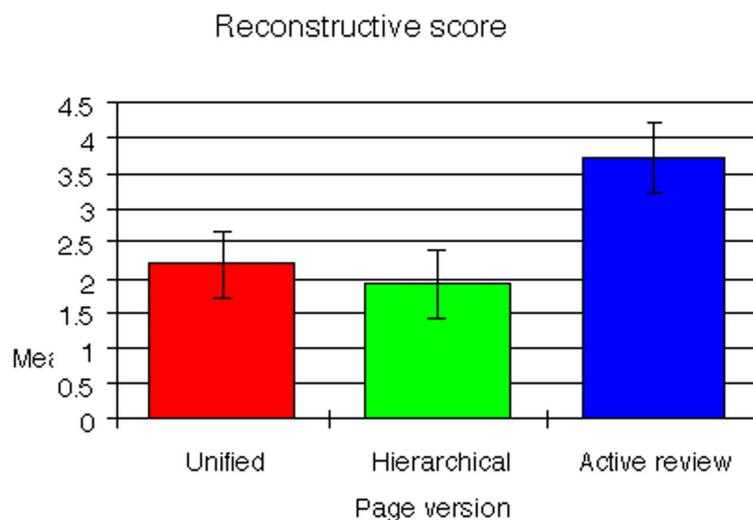
Figure 2: *Graph of scores on the reconstructive questions for each web structure*

Document structure	Mean score	Standard deviation
Unified	2.187	1.974
Hierarchical	1.929	1.940
Active review	3.750	1.422

Table 3: *Mean scores and standard deviations on the reconstructive questions for each web structure*

A one-way ANOVA test showed there was a significant difference between groups $(F(2,41)=3.73, p=0.033)$. The active review group scored significantly more than the hierarchical group $(t(23)=2.75, p=0.0057,$ one-tailed) and the unified group $(t(26)=-2.33, p=0.01415)$. There was no significant difference, however, between the unified and hierarchical versions $(t(27)=-.36, p=0.36,$ one-tailed).

3.   Experiment 2: The browsing effect

In order to separate the effects of revision questions from page size, this second experiment directly compares a unified and a hierarchical hypertext.

3.1   Method[7]

28 students from a first-year 'Statistics for Psychology' course took part in this experiment as one of their class assignments. They had little previous statistical training, having no post-16 mathematics qualifications, apart from this course. All had received instruction and practical experience in using a World Wide Web browser.

The participants were randomly assigned to one of two groups corresponding to a hierarchical or unified hypertext. The hierarchical hypertext[8] consisted of a contents page leading to a set of 13 pages. The unified version hypertext[9] contained the same information in one page. This information was roughly twice as long as that given in *Expt.1: Comparing three common document structures*. The hypertexts gave information on non-parametric tests, a new topic

---

[7]   *This experiment improves on several methodological points in Comparing three common document structures. Longer test materials are used as part of a class assignment, to better approximate a real learning situation. Thanks to Simon Buckingham Shum for suggesting them.*

[8]   *http://www-jime.open.ac.uk/98/12/demos/stats/stats2.html*

[9]   *http://www-jime.open.ac.uk/98/12/demos/stats/stats.html*

halfway through the statistics course. While participants may have differed in mathematical ability or previous learning, their random distribution between groups should prevent this influencing the results.

Netscape was set up at the correct starting page on a cluster of 486 Windows-based PCs. Each student was instructed to study their hypertext for 10 minutes, ready to take a 5-minute paper-based test afterwards. The test contained 10 reconstructive and 6 reproductive randomly-ordered questions. A handout was prepared for the following week's class explaining the results and thanking the students for their participation. A pointer to an e-mail address and WWW page was given for any further queries.

3.2   Results

Participants' scores[10], in total and separated into reproductive and reconstructive components, are shown below.

Test scores
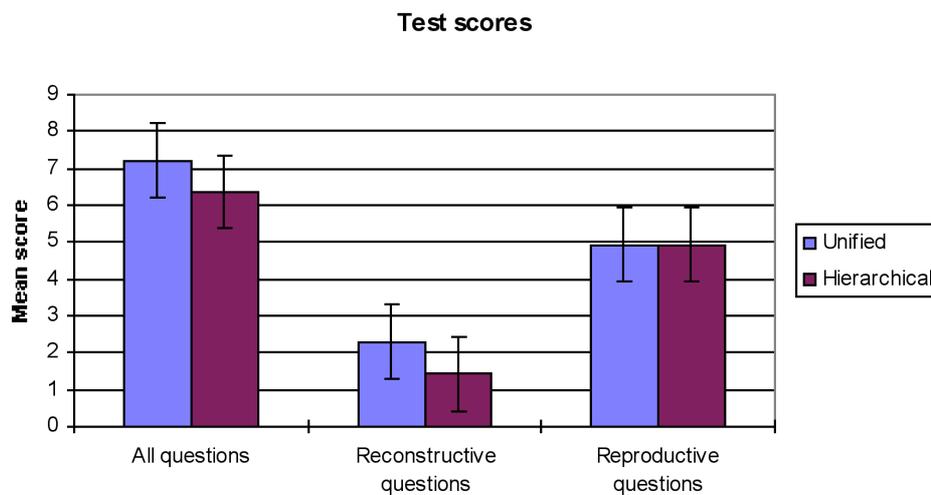
Figure 3: *Graph of scores in Experiment 2*

Document structure	Total score		Reconstructive score		Reproductive score	
	Mean	SD	Mean	SD	Mean	SD
Unified (n=14)	7.21	3.51	2.29	2.16	4.93	2.16
Hierarchical (n=14)	6.36	3.39	1.43	1.55	4.93	2.64

Table 4: *Mean scores and standard deviations in Experiment 2*

[10]   *Again, scores were corrected for guessing, on questions 2, 6(i), 7 and 10(i)*

A series of analyses showed there were no significant differences between the scores of participants reading the unified hypertext and those reading the hierarchical hypertext for total score (t(26)=0.66, p=0.26, one-tailed), reproductive score (t(26)=0.00, p=0.50, one-tailed) or reconstructive score (t(26)=1.34, p=0.096, one-tailed).

4.   Discussion

WWW document structure does seem to have an effect on information retrieval. The different structures in *Comparing three common document structures* led to a difference of a full standard deviation between the highest and lowest recall documents. Because reconstructive and reproductive learning were affected differently, there was no significant difference between total recall scores for the three document types.

The low ranking of the hierarchical hypertext for reconstructive learning seems to confirm the worries of Jonassen (1993) and Dillon et al. (1993) about the concept of semantic structures. As Hammond (1993) states, "there are many situations where learning is most effective when the freedom of the learner is restricted to a relevant and helpful subset of activities." (p52)

Dillon et al. make the important point that paper texts are also structured. Might they not be just as successful in conveying semantic information? Few documents consist of page after page of uninterrupted text. Section headings and bullet points are just two of the devices authors use to give structural information. The restructuring of a document that often occurs when it is moved on-line may be more important than any effect the medium itself has.

In addition, 'standard' structures such as the newspaper article - or an APA style report - are more familiar to readers than new structures created by hypertext designers. Kintsch and Yarborough (1982) showed that standard text structures prompt better understanding than unconforming texts. Hypertext design tools, and the whole 'flash-bang' culture of the Internet, promote novelty as a positive attribute for WWW pages. For learning materials, this may be a mistake.

Even if hypertext structures do impart useful semantic information, "such high-level abstractions are always going to be in danger of 'spoon-feeding' students with structures they should be developing for themselves." (Whalley, 1993, p14). At university level, the development of such 'thinking skills' is perhaps even more important than the acquisition of knowledge (Bligh, 1977). Using revision questions in the text appears to be a better way of stimulating semantic learning. It is important, however, that they encourage rather than replace original thinking by the student.

Conversely, it appears that splitting pages into small pieces has no detrimental effects on later information retrieval. Experiment 2 found that readers jumping between small sections of text recalled as much as those reading a long page of material. Nash's (1980) expository flow therefore seems unimportant in this context. It is a moot point whether clicking on a link, scrolling down a screen or turning the page of a book breaks concentration more. At least for short-term recall, the medium seems unimportant to the message.

The varying performance of the active review hypertext in *Comparing three common document structures* is initially puzzling. Samarapungavan and Beishuizen (1990) also found that encouraging students to take different perspectives on hypertexts improved what they called conceptual learning, but not factual recall.

Work by Neff Walker (1986) explains this difference. He showed that elaboration of memory traces actually *decreases* recall of factual information, by reducing the strength of pathways between the original proposition and other stored data. This reduces the likelihood of activation and thus recall of the original fact, which cannot be inferred from the elaborated information stored with it (unlike reconstructed facts). The elaboration caused by the revision questions in the active review hypertext was most likely responsible for participants' poor performance on reproductive recall. This is further evidence against Anderson and Reder's (1979) contention that elaboration increases the likelihood of activation of propositions adjacent to the original memory trace, and thus the probability of its recall.

Walker, Jones and Mar (1983) suggest that the increased amount of cognitive effort caused by the revision questions would improve the recall of that information by increasing the 'distinctiveness' of the memory trace. This approach has been criticised by Mitchell and Hunt (1989), who found in the literature a 'haphazard correlation between indexes of cognitive effort and of memory performance'. Semantic processing sometimes requires more cognitive effort, but not always (Eysenck and Eysenck, 1979). *Comparing three common document structures* appears to back up their conclusion that cognitive effort serves only as a boundary condition in memory performance - any extra processing caused by the revision questions resulted in diametrically opposed changes in recall performance for reconstructive and reproductive learning. The explanation that cognitive effort acts in the same manner as elaboration, in opposite directions for the two types of learning, would only be viable if creating elaborated semantic memories was a less costly process than storing facts. This would fly in the face of 20 years of cognitive science research.

Participants' cognitive resources were anyway unlikely to have been stretched to breaking point by the simple task of learning written material - something at which undergraduates will have had plenty of practice. Mitchell and Hunt propose that only when other unrelated cognitive

processes 'crowd out' memory formation processes will high cognitive effort cause memory systems to fail. Changes in cognitive effort may be sufficient to produce changes in recall performance, but are not a pre-requisite.

It may also be that, in the five minutes available to participants in the first experiment, a trade-off had to be made between learning facts - where organisation is perhaps more important - and creating elaborated, inferential memories. The cueing of reconstructive learning caused by the revision questions in the active review hypertext may have distracted participants from reproductive learning. Memory is actually a good index of task interference (Mitchell and Hunt, 1989). As Walker (1986) says, "processes that rely upon direct retrieval do not play an important role in explaining the generally beneficial effect that elaborative processing has upon recall" (p.325). A Mitchell and Hunt-style boundary condition may have caused the processes initiated by the revision questions to prevent other memory processes completing successfully. This is the most common reason why memory fails.

It appears therefore that the revision questions were the key to the first experiments' results. The second experiment found no evidence of any browsing effect. There was a full standard deviation's difference between the active review hypertext and its two competitors in *Comparing three common document structures*, positively for reconstructive learning and negatively for reproductive learning. It is ironic, although reassuring, that 'traditional' memory theory has proven more important in explaining this experiment's results than any new computing-specific theories. In return, this experiment has provided extra evidence against the hypotheses of Anderson and Reder (1979), and for those of Walker (1986).

The revision questions within the active review hypertext required a mix of reproductive and reconstructive answers. The test materials were not long enough to extract meaningful data on which type of question was more successful in encouraging either type of learning. This may be an important point for further work, with particular practical application for authors of learning materials (Whalley, personal communication). Some previous work has indicated that semantic tasks produce better recall that non-semantic (Tyler, Hertel, McCallum and Ellis, 1979; Krinsky & Nelson, 1981). Mitchell and Hunt (1989) caution, however, that extra tasks must stimulate new memory processes above those caused by the original task to improve recall. Revision questions, especially those which encourage readers to consider the information given in new ways, would seem to fulfil this requirement. More complex question types, particularly where readers must assess their confidence in their answers, could encourage this process further. The active review hypertext has the advantage that students are forced to answer the questions before proceeding, unlike paper texts!

Like the unified hypertext, it also ensures readers are lead in the correct order through the

material. The hierarchical version allows users to choose their own route through the text. This may be inappropriate for learning materials, where an expert author attempts to convey more than simple snippets of knowledge. In larger hypertexts it may also cause navigation problems, with users missing material or becoming 'lost in hyperspace' (Edwards and Hardman, 1989). WWW servers maintain accurate logs of page accesses and times, which would allow a precise picture of users' paths through a hypertext to be built up. This would overcome the problem with navigation research identified by Dillon (1992), of measuring the reading process as opposed to the reading outcome.

Perhaps the two texts were not long enough for the unified structure's other advantages to show through. It is difficult to develop a complex rhetorical argument in less than a thousand words. The test questions measuring reconstructive learning required participants to integrate material usually from only one paragraph. It would be interesting to investigate how a broken-up hypertext compares with a unified text on questions which evaluate longer sections of text, over such section breaks.

Hypertext should not be completely written off at this point. This experiment considered structure, which is only one small variable. The materials used were simple combinations of text and images. The real potential of computer-aided learning lies in richer, more interactive materials. Recent work along these lines (e.g. Large, Beheshti, Breuleux and Renaud, 1994) has found that learning is often improved by taking advantage of techniques such as animation which are not possible in print. Dynamic interaction between learner and computer is even more important, particularly in the opportunity for individualised attention that systems can give each student. Communication between students while reading hypertexts takes this a step further, opening up many possibilities for peer-group learning. Asking the students themselves to write the materials as a collaborative effort provokes far more active learning than simply spoon-feeding them reams of materials (Downing and Brown, 1997).

Further work with class-based, longer and richer hypertexts would therefore be the best extension to this experiment. It would allow proper evaluation of the unified hypertext, taking into account factors such as the reinforcement of concepts and themes which Whalley (1993) considers important. Participants would use more realistic strategies for learning the material than those available in five or ten minutes. Other computer media could be included in the comparison. Most importantly, it would best approximate real-life learning situations. Ultimately, improving educational practice must always be the goal of this type of research.

Acknowledgements

References

Anderson, J.R. and Reder, L.M. (1979). *An elaborative processing explanation of depth of processing*. In L. Cermak and F. Craik (Eds.), *Levels of processing in human memory*. Hillsdale, NJ: Erlbaum.

Bligh, D.M. (1977). *Are teaching innovations in post-secondary education irrelevant?* In M.J.A. Howe (Ed.) *Adult Learning: Psychological Research and Applications*, p249-266. London: Wiley.

Bush, V. (1945). *As we may think*. Atlantic Monthly, 176/1, 101-108.

Collins, A.M. and Loftus, E.F. (1975). *A spreading activation theory of semantic processing*. Psychological Review, 82, 407-428.

Collins, A.M. and Quillian, M.R. (1969). *Retrieval time from semantic memory*. Journal of Verbal Learning and Verbal Behaviour, 8, 240-247.

Craik, F.I.M. and Lockhart, R.S. (1972). *Levels of processing: a framework for memory research*. Journal of Verbal Learning and Verbal Behaviour, 11, 671-684.

Dillon, A. (1992). *Reading from paper versus screens: a critical review of the empirical literature*. Ergonomics: 3rd Special Issue on Cognitive Ergonomics, 35(10), 1297-1326.

Dillon, A., McKnight, C. and Richardson, J. (1993). *Why Physical Representations are not Semantic Intentions*. In McKnight, Dillon and Richardson (1993) p169-191

Downing, A.C. and Brown, I. (1997). *Learning by cooperative publishing on the World Wide Web*. Active Learning, 7, 14-16.

Edwards, D. and Hardman, L. (1989). *"Lost in hyperspace": cognitive mapping and navigation in a hypertext environment*. In R. McAleese (ed.) *Hypertext: Theory into Practice*. Oxford: Intellect. p105-125.

Eysenck, M.W. and Eysenck, M.C. (1979). *Processing depth, elaboration of encoding, memory store, and expended processing capacity*. Journal of Experimental Psychology: Human Learning and Memory, 5, 472-484.

Fiderio, J. (1988). *A grand vision*. Byte, October, 237-243.

Hammond, N. (1993). *Learning with Hypertext: Problems, Principles and Prospect*s. In McKnight, Dillon and Richardson (1993) p51-69

Jonassen, D.H. (1993). *Effects of Semantically Structured Hypertext Knowledge Bases on Users Knowledge Structures*. In McKnight, Dillon and Richardson (1993) p153-168

Kintsch, W. and Yarborough, J. (1982). *The role of rhetorical structure in text comprehension*. Journal of Educational Psychology, 74, 828-834.

Krinsky, R. And Nelson, T.O. (1981). *Task difficulty and pupillary dilation during incidental learning*. Journal of Experimental Psychology: Human Learning and Memory, 7, 293-298.

Large, A., Beheshti, J., Breleux, A. and Renaud, A. (1994). *Multimedia and Comprehension: A Cognitive Study*. Journal of the American Society for Information Science. 45(7): 515-528.

McKnight, C., Dillon, A. and Richardson, J. (Eds.) (1993). *Hypertext: a psychological perspective*. Chichester: Prentice Hall.

Mitchell, D.B. and Hunt, R.R. (1989). *How much effort should be devoted to memory*. Memory and Cognition, 17(3), 337-348.

Nash, W. (1980). *Designs in Prose*. London: Longman.

Samarapungavan, A. and Beishuizen, J.J. (1990). *Hypermedia and knowledge acquisition: learning from non-linear expository text*. Presented at the EARLI "Text Processing' Special Interest Group, University of Amsterdam, November 8-9. Quoted in Whalley (1993). Tyler, S.W., Hertel, P.T., McCallum, M.C. and Ellis, H.C. (1979). *Cognitive effort and memory*. Journal of Experimental Psychology: Human Learning and Memory, 5, 607-617.

Walker, N. (1986). *Direct retrieval from elaborated memory traces*. Memory and Cognition, 14(4), 321-328.

Walker, N., Jones, P. and Marr, H.H. (1983). *Encoding processes and the recall of text.* Memory and Cognition, 11(3), 275-282.

Whalley, P. (1993). *An Alternative Rhetoric for Hypertext.* In McKnight, Dillon and Richardson (1993) p7-17

Wickelgren, W.A. (1977). *Learning and memory*, p231-232. New Jersey: Prentice-Hall.

Appendix A: Test sheets for Experiments 1 and 2

Expt 1: Miro Questions

These were the (randomly sorted) questions participants attempted after reading the Web page(s) they were assigned to. The letter at the start indicates whether the question tested reconstructive (C) or reproductive (P) learning.

1. P Which Manifesto influenced Miro?
2. P Who published it?
3. P What was the name of the Academy Miro entered?
4. P Which distinguished artist did he meet there?
5. C Did Miro want to live the stereotypical 'poor' artists' life?
6. C Was Miro's father worried while his son was at the academy?
7. C Was Miro's mother an emotional woman?
8. C What changed Miro's early preference for naturalistic art?
9. P Which city was Miro born in?
10. P In which year did Miro have his first exhibition?
11. P Name the two pictures by Miro shown and mentioned in the text
12. C Was Miro a supporter of Spanish unity?
13. C Did Miro respect his contemporaries in Paris?
14. P Who commissioned the Sun and Moon walls?
15. P Which award did they lead to?
16. C Did Miro's parents think art was a good career?
17. P Which two unusual materials did Miro favour?
18. P Which famous artist befriended Miro in Paris?
19. P Which art movement first influenced Miro?
20. P Who was their most famous member?

Expt 2: Statistics Questions

These were the (randomly sorted) questions participants attempted after reading the Web page(s) they were assigned to. The letter at the start indicates whether the question tested reconstructive (C) or reproductive (P) learning.

1. P Why are rank-based tests more powerful than sign tests?
2. C Are parametric or non-parametric tests better for uncovering effects?
3. P Which test uses a statistic U?
4. P How do you deal with an observed value of zero in a signed rank test?
5. P Which test can be used as a non-parametric alternative to the one-sample and paired t-test?
6. C If a Wilcoxon Rank Sum test found a significant effect in a set of data, would a Mann-Whitney also find this effect? Why?
7. P Are non-parametric tests more or less complex than parametric ones?
8. P What 3 assumptions do parametric tests rely on?
9. P What are the two non-parametric alternatives to the two-sample t-test called?
10. C If you were analysing a lot of data by hand looking for a large effect, would you use sign or rank-based tests? Why?