

Effective Illustrations in Interactive Media: What Works?

Barbee Teasley

Human Factors
Ameritech

2000 W. Ameritech Center Drive
Hoffman Estates, IL 60196 USA
01-847-248-1400
FAX 01-847-248-6043
e-mail: teasley@acm.org

**Keith Instone, Laura Marie
Leventhal and Eric Brown**

Computer Science Department
Bowling Green State University
Bowling Green, OH 43403 USA
01-419-372-2765
FAX 01-419-372-8061
e-mail: instone@acm.org

ABSTRACT Illustrations are often used in interactive media with the goal of enhancing information retrieval and learning. Yet prior work has shown that illustrations on a screen are not necessarily used effectively, and may not even be used at all. Two studies examined factors related to the successful use of illustrations. Study 1 contrasted three interaction styles. In one style, only text appeared when a term was selected. The second presented text as a caption, while the third revealed the text next to the relevant portion of the illustration. The last style produced superior learning of both illustration-related and unrelated material. The caption condition was no better than the control, no-illustration condition. Study 2 showed learning is just as effective if the user selects from a list and reads near the illustration, or selects from the illustration and reads a caption, suggesting that the most important factor is compelling users to direct their attention to the illustration.

KEY WORDS illustrations, interactive media, multimedia, learning

1. INTRODUCTION

Illustrations are ubiquitous in interactive media such as web sites, CD-ROMs and computer-based instruction. They can be used for many purposes, ranging from purely decorative, to supplemental, to crucial information (Waddill and McDaniel, 1992). While graphic design principles provide guidelines on the aesthetic aspects of illustrations and layout, there is little guidance available on how to make illustrations effective as a means of transmitting information which enhances or supplements information presented in text. In the absence of such rules, designers may use intuition—a practice which may or may not be effective (Teasley, Leventhal, Blumenthal, Instone and Stone, 1993). Furthermore, there is no clear understanding of the general cognitive principles which underlie learning from interactive, illustrated text.

Given the well-known phrase that "a picture is worth a thousand words," most would intuit that a computer-based presentation of information including illustrations would be easier to learn from and seek information in than text-only forms. In some situations, this is certainly true. For example, Ogozalek (1994) showed that elderly users performed better on information retrieval tasks using either of two computer-based forms of a document than a printed paper document. The computer-based version containing video, graphics, and audio, as well as text, produced better performance than the text-only computer version. The multimedia presentation was also the most highly preferred. The most commonly cited reason was that it overcame vision problems associated with aging. Similarly, Harrison (1995) found that online help incorporating either still or animated graphics improved performance on procedural tasks, compared to text-only help.

However, a surprising number of other studies have found that simply integrating illustrations and text on a computer screen does not, ipso facto, increase performance. For example, Wright, Hull and Black (1990) looked at subjects' willingness to study diagrams presented as an adjunct to text and at the amount of information gained from the diagrams. They found that when subjects had free access to diagrams (e.g., a button present on the screen labeled "Diagram"), they spent relatively little time studying the diagrams and made minimal use of the information in the diagrams. This was true despite variations in the diagrams and in the instructions. However, when compelled to look at diagrams because they came up unexpectedly, performance improved. This suggests that simply adding a button or notation to on-screen text to signal that additional information is available may not be a truly effective way to enhance learning.

Glenberg and Kruley (1992) investigated the impact that pictures presented with text on a computer had on the resolution of anaphors. An anaphor is a word or phrase that is interpreted in relation to previous elements in a discourse. An example of an anaphor is using a pronoun, such as: "The insect had spots. *It* was a ladybug." Earlier work had indicated that the greater the distance separating the antecedent (the initial reference) and the anaphor, the poorer the comprehension. In the study, subjects read texts that contained short descriptions of objects. Four or five parts of the objects were described in the text. In the text, antecedents described the name and location of the part. The anaphor repeated the spatial location of the part described in the antecedent, and also elaborated on its function. One-third of the texts had no pictures, one-third had an accompanying picture that was always present, and one-third had a picture that either disappeared or appeared late. They found that the presence of pictures did generally help, but did not differentially improve comprehension of texts requiring far anaphor resolution. "We are still somewhat puzzled. The pictures did not assist anaphor resolution, but there is no logical reason why they should not have done so" (Glenberg and Kruley, 1992, p. 471). These findings suggest that subjects did not use the information in the picture effectively to resolve the anaphors. The authors suggest that using illustrations to improve reading comprehension is a skill that must be learned. It could also be the case that

illustrations which merely accompany the text in a passive way do not improve comprehension.

A study of hypermedia usability compared question-answering performance using a hypermedia encyclopedia to an equivalent paper version (Leventhal, Teasley, Instone, Rohlman and Farhat, 1993). While users were quite successful in most tasks with the hypermedia encyclopedia, they were not particularly effective in using illustrations as a source of information. In a later study, two hypermedia versions of the same encyclopedia were compared (Instone, Teasley and Leventhal, 1993). One version (the one used in the earlier study) included overlapping windows and several alternative ways to accomplish the same task. The other version used tiled windows and had simplified tools. While the content was identical in both cases, users were more successful at using illustrations in the new, simpler version (and were also much better at using other tools, performing better than paper book users in many instances). The results offered no definitive explanation of this unexpected result. It is possible that improving the interface as a whole improved users' ability to comprehend the text and illustrations, due to a reduced cognitive load. That is, fewer cognitive tasks may have been necessary to use the tools and attain access to information, allowing the user to spend more effort on remembering the content itself.

A study by Egan, Lesk, Ketchum, Lochbaum, Remde, Littman and Landauer (1991) found that chemists had great difficulty performing a task called "analogous transformation" while using a hypermedia/document browsing version of a collection of journal articles. The analogous transformations task requires that the user access a number of illustrations and draw inferences based on the illustrations. In the hypermedia, the user had to first find the appropriate article, and then perform an action to pop up the illustration. Subjects had problems both in coding the pictorial representation of the problem correctly (causing them not to find the appropriate article), and in failing to pop up the illustration, even when they did access the correct article. The authors note that "given that information linked to a text is displayed only at the option of the user, how should that information 'make itself known' to the user?...Having linked information call attention to itself at appropriate times is an unsolved and difficult problem" (Egan, et al., 1991, p. 310). These

findings, similar to those of Instone, Brown, Leventhal and Teasley (1993), suggest that illustrations which are not well integrated with the text will be poorly utilized.

Finally, a series of studies looked at readers' willingness to use a glossary containing the definitions of words used in a text, and which also included pictures of the terms being defined (Black, Wright, Black and Norman, 1992). The studies showed that people were more inclined to use the glossary when the terms in the text were marked by a superscript black dot than when they were unmarked. They also found that moving the list of available glossary terms to the left side of the screen and requiring the user to access the items through that list reduced readers' willingness to access the items and their associated pictures. The authors suggest that there are cognitive costs associated with changing modality (e.g., switching from text to illustrations) and that this cost impacts on the processing of the main text.

These studies all show that it is not sufficient to simply add relevant illustrations to text and expect them to automatically aid comprehension. In particular, it appears that the amount of effort required to choose and access the illustration influences its use. Having the user perform an added action such as clicking on a button, or requiring the user to refer to information presented peripherally (e.g., from a list off to the side) deters the effective use of illustrations. Even passive display of an illustration (i.e., the Glenberg and Kruley result) does not insure its use. What did appear to encourage integration of the information from both modalities was the unexpected appearance of an illustration. This approach requires no action from the user, and uses a centralized (not peripheral) display of the information. Unfortunately, while this technique does enhance learning, it violates a central principle of user interface design by not allowing the user to control the interaction. We wished to explore alternative ways to involve the user with an illustration, without taking control away from the user.

The goal of the present studies was to explore factors which influence the effectiveness of integrating text and illustrations. We hypothesized that the way a user interacts with illustrations influences their effectiveness, and that an interaction style which essentially compels the user to interact with an illustration would be most effective. To that end, the first study contrasted text-only and traditional caption-

style presentation of an illustration, to a style in which the text was visually next to the relevant portion of the illustration. As a control, all three conditions required the user to perform a point-and-click action to expose the text.

2. STUDY 1

The hypothesis of the first study was that compelling the user to interact with an illustration would result in enhanced learning through more integration of the textual and graphically presented information. The means chosen for compelling interaction was to present the illustration first, and to reveal textual information directly adjacent to the relevant portion of the illustration. The illustration contained "hot spots" (areas that could be clicked on with a mouse) that, when activated, revealed the relevant text. We labeled this an *integrated* interaction style. (See Figure 2.)

Two other interaction styles served as control conditions. One was a text-only condition. In this case, there were hot spots on the screen which were small circles next to text terms. When the user clicked on a hot spot, the text appeared directly to the right. The second control condition, the caption style, was exactly like the text-only condition, except that an illustration was present in the area above the hot spots and text (Figure 1). This style mimics a typical paper publication, with a captioned illustration. Note that all three conditions require the user to use the mouse and cursor to click on a hot spot and activate the text. Thus, the use of mouse and cursor is constant across all three conditions, as is the need to perform an action to reveal the pertinent text, providing control for these factors. By controlling for use of a mouse and cursor, and by always requiring the user to perform an action to see the text, we hoped to isolate just the effects of 1) presence or absence of an illustration, and 2) captioning versus integrated text and illustration.

We hypothesized that the subjects would learn the best when the text was integrated with the illustration. We also expected that having the text below the illustration would be no more effective than text alone, due to the relative isolation of the text from the graphic.

2.1 Materials

The materials consisted of a series of topic screens. Each topic was related to incidents from the Sherlock Holmes detective stories, by Sir Arthur Conan Doyle, and contained information about geographic entities or devices of interest to Sherlock Holmes. Sample topics are: the reproductive system of spaghnum sporophyte, areas near Ireland, obscure names for parts of a horse, and sections of Scotland. For each topic, there were four or five items. Each item consisted of one piece of spatial (directional) information and one fact related to a Sherlock Holmes story or about the object. For example, for the topic of Switzerland, one item was the town of Lausanne. The spatial information about Lausanne was "A town in west Switzerland." The fact was "Lady Frances Carfax was last heard from here."

The topics were presented in three interaction styles: *text-only*, *caption* and *integrated*. In the text-only interaction style, there was no illustration. Four or five item names were listed at the bottom of the screen. When users moved the cursor over the item name, text information about that item popped up to the right of the name. The text appeared in the same location on the screen for each item. (The text-only style was identical to what is shown in Figure 1, except there was no illustration. Note that in the

figure, the user has selected "Davos Platz" and the text has appeared to the right).

The caption interaction style was similar to the text-only; except that an illustration (map or picture) appeared above the list of items, as shown in Figure 1. As in the text-only style, the text appeared to the right of the list of items. Each item name also appeared in the illustration in bold text.

In the integrated style, users pointed directly at the item on the illustration to reveal the text information. As shown in Figure 2, the information appeared near the location of the relevant item on the illustration.

In order to vary the materials, not every topic was geography-based. Several of the topics dealt with objects. For example, one topic showed the different parts of a gasogene, a device Holmes used for making soda water.

Each subject saw 12 topics, in a within-subjects design. The three interaction styles were randomly distributed across the topics, with the limitation that each user saw four topics in each interaction style. The order of the topics was also randomly generated.

Following the presentation of each topic, users saw a screen which directed them to work on a word search task (on paper) for 30 seconds. The word search task served as a distracter, so that the users could not answer the subsequent questions based on short term memory. At the end of 30 seconds, they answered a series of eight multiple choice questions presented on the computer. Four of the questions concerned the fact information from the topic and four concerned the spatial information. All eight questions were the same for each topic, regardless of interaction style used to present the topic. In addition to scores on the questions, the time spent studying each topic was recorded.

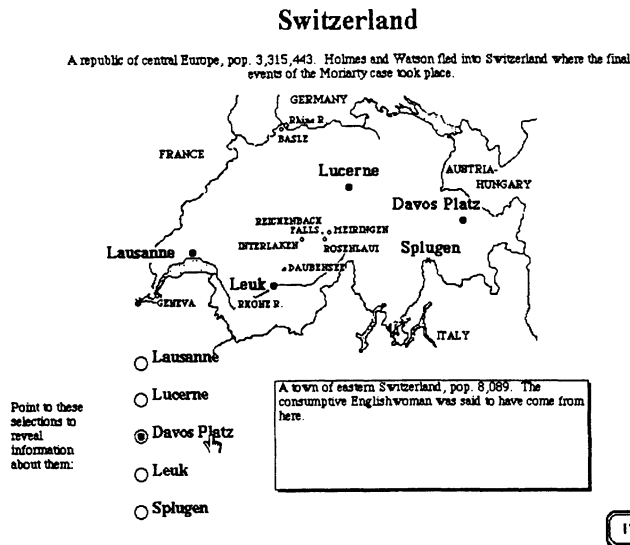


Figure 1: The caption interaction style

2.2 Subjects

The subjects were 27 students from an introductory computer science class. All had previous experience using a mouse for pointing. All rated themselves as "not very familiar" with the Holmes stories.

2.3 Procedure

The subjects were first given training for the task by viewing a sample topic and demonstrations of the three interaction styles. In these examples, instructions appeared on the screen along with the text and illustrations. The instructions explained where to point and what button to press when they were finished with each topic. Next, a sample topic was presented in one of the three interaction styles (randomly chosen), with no instructions present. This initial topic was the same for all users, and served as a practice trial. They then performed the primary task, which consisted of 12 trials. Each trial included the presentation of the topic (which the users were allowed to study as long as they

wished), followed by 30 seconds of distracter task, followed by the eight multiple choice questions. Time for answering the questions was not limited.

2.4 Results

For each topic, there were a total of four spatial information questions and four factual information questions. A score of one was given for each correct answer and a zero for each incorrect answer. There were four topics presented in each of the three interaction styles. Thus, the maximum score for each type of question (spatial or fact) for each of the three interaction styles was 16. The mean number correct (out of 16) for each condition is shown in Table 1.

	Interaction Style			
	Text-only	Caption	Integrated	Overall
Spatial	8.6	9.9	11.3	9.9
Fact	10.2	11.0	12.3	11.1
Overall	9.4	10.4	11.8	

Table 1: Mean scores on the multiple choice questions in Study 1 as a function of type of question (spatial vs. fact) and interaction style. $N = 27$.

Using a repeated measures ANOVA, the main effects of question type and interaction type were significant ($F(1,34) = 16.1, p < .001$ and $F(2,34) = 9.7, p < .001$, respectively). The mean total correct was higher for the fact questions (mean = 11.1) than for the spatial questions (mean = 9.9). This suggests that the spatial information was more difficult to learn overall than the factual information, in spite of the fact that for two-thirds of the topics, the illustration was present and might have been expected to serve as an aid to learning the material. The factual information, on the other hand, would not be particularly enhanced by the illustration (e.g., the fact that "Lady Frances Carfax was last heard from here" is not actually illustrated or demonstrated by the illustration). Overall, the total number of correct answers for the integrated style was higher than for either of the other two styles (mean = 11.8 versus 10.4 for caption and 9.4 for text-only). This suggests that the integrated style is the most effective way to present either factual or spatial information.

To explore whether the presence of any illustration was useful, two sets of pairwise t-tests were performed. One was done on the spatial data, and

another on the fact data. Comparisons of the spatial question data indicated that there was no significant difference between the text-only and caption conditions, indicating that the presence of the illustration did *not* enhance learning of spatial information when the textual material was presented as a caption. In addition, scores on spatial questions for the integrated interaction style were significantly higher than the caption style (paired t, two-tailed = 2.0, $DF = 26, p < .05$). This shows that the critical factor which enhanced the learning of the spatial material was the integration of the text with the illustration, and not solely the presence of the illustration.

A second set of pairwise t-tests was done on the fact questions. The results parallel those found for the spatial questions. For the fact questions, the text-only and caption styles were again not significantly different from each other, suggesting that the simple presence of an illustration did not enhance learning. On the other hand, the comparison of the caption scores to the integrated style scores was significant ($t = 2.4, DF = 26, p < .05$), indicating that the integrated style, and not just the presence of the illustration, was responsible for the improved learning.

The time spent studying each topic card was also analyzed. The mean times were 70.5 seconds for text-only, 73.7 seconds for caption and 67.6 seconds for integrated. These differences were not significant.

2.5 Summary of Study 1

The results of Study 1 show that an integrated style is the most effective way to present text and illustrations when the user wishes to remember the information presented. Furthermore, the presence of an illustration is of no added benefit unless the text and illustration are integrated.

Surprisingly, enhanced learning occurred for both factual and spatial information. One might assume that the integrated style would be more of an aid to memory, because the text is presented in a spatially-related way, thus reinforcing the information in the illustration. There is no *a priori* reason why the integrated style would aid recall of the factual information, which was not represented by the illustration. The results, however, show that learning of both sorts of information was enhanced. Perhaps this stems from the fact that the integrated style helped the reader learn the organizational framework inherently present in the illustration, and provided "hooks" for the factual information. The time data, which was not different for the three styles, indicates that the advantages of the integrated style were realized with no additional time requirements.

3. STUDY 2

Study 1 demonstrated that integration is superior to either text alone or an illustration with a caption for learning both factual and spatial information. We then speculated about exactly what aspect of the integrated style contributes to this enhancement. Considering the actions of people when they use the integrated style suggests that they actually do two things:

- they point to the relevant part of the illustration to activate the related text, and
- they read the text which appears near the relevant portion of the illustration. (e.g., moving the hand in a "north" direction to click on a city located in the north on a map). This is similar to the way clapping a rhythm with your hands might reinforce hearing a rhythm with your ears. A second question is whether the placement of the text is critical. When the user reads the text, is it important that the text be next to the relevant portion

of the illustration? Perhaps readers can be seeing, at least peripherally, the illustration at the same time they are reading, thus getting more information. Perhaps learning is facilitated because readers do not have to shift their eyes up and down between the illustration and the text. Or perhaps it requires a combination of these processes to achieve the enhanced learning. Study 2 attempted to answer these questions by comparing three somewhat different styles of interaction, all of which used an illustration, but involved different ways of revealing the text.

The content of the instrument was the same as in Study 1, including both factual and spatial questions. Three styles of interaction were used. One was the integrated style from Study 1. That is, the user pointed to an item name on the illustration to reveal the text, which appeared nearby. A second interaction style, called *illustration point*, had the user point to the item name on the illustration, while the text appeared as a caption. The purpose of this condition was to see if the action of moving the cursor onto the relevant part of the illustration was more important than having the text appear nearby. In the third style, called *illustration read*, the item names appeared in a list underneath the illustration. When the user moved the cursor over an item name, the related text appeared next to the relevant part of the illustration. The purpose of this condition was to see if having the text next to the illustration was more important than moving the cursor onto the relevant part of the illustration. The population of subjects for Study 2 was identical to that of Study 1. There were 20 participants.

3.1 Results

The mean number of questions correct for each of the conditions is presented in Table 2. (Recall that the maximum score for any condition is 16.) As in Study 1, performance was better on the fact questions than the spatial questions ($F(1,38) = 17.6, p < .001$). No other differences were significant. As Table 3 shows, the three integrated styles (point, integrated and read) clearly emerge as the better strategy for presenting text and illustrations. Comparing the characteristics of these three styles to the caption and text-only conditions, a common feature of the integrated styles is that users must at some time during the process move at least their eyes (and possibly the cursor) onto or near the relevant portion of the illustration. In the integrated style, the user moves the cursor onto the

item name in the illustration, and reads the text near the illustration. In the illustration point style, the user moves the cursor onto the item name in the illustration, and in the illustration read style, the user reads the text near the illustration. In contrast, with the caption style, the user need never look at the

illustration. (Of course, there was no illustration in the text-only style.) This suggests that the critical factor for enhanced learning is that the user must be somehow compelled to interact visually with the illustration at some point in the study period.

	Interaction Style			
	Integrated	Illustration Read	Illustration Point	Overall
Spatial	11.8	11.5	11.6	11.6
Fact	13.6	12.1	12.9	12.8
Overall	12.7	11.8	12.2	

Table 2: Mean scores on the multiple choice questions in Study 2 as a function of type of question (spatial vs. fact) and interaction style. N = 20

Interaction Style	N	Fact	Spatial	Mean
Illustration Point	20	12.8	11.6	12.2
Integrated	47	12.8	11.5	12.1
Illustration Read	20	12.1	11.5	11.8
Caption	27	10.9	9.8	10.4
Text-only	27	10.2	8.6	9.4

Table 3: Combined results from Study 1 and Study 2.

3.2 Discussion of Study 2

The results of Study 2 indicate that neither the movement of the cursor nor the placement of the text is more important to the enhanced learning from the integrated style. However, consider the results of both studies together, as shown in Table 3. (For the integrated style, the mean score combined across the two studies was used. The difference between these two means was not significant.)

4. CONCLUSIONS

One of the exciting aspects of interactive media is that they can provide a more efficient and effective way of presenting information and of promoting learning than through the use of more traditional media. However, prior studies have found that illustrated information presented on a computer display is not always used effectively. In fact, an illustration can be of no additional value at all, compared to plain text. In order to create effective presentations, designers need

an understanding of the factors that make illustrations effective.

One factor which had been shown in a prior study to enhance learning from illustrations was compelling users to look by having the illustrations appear unexpectedly (Write, Hull and Black, 1990). However, this approach violates usability principles, and may not be suitable in many situations. In what we call the integrated style of interaction, the user must actively select the portion of the illustration of interest in order to reveal the textual information, which then appears adjacent to the illustration. This turns the user from a passive reader of text who has the option of looking at an illustration (the style typical of paper media), to an active user, who must look at and process the illustration to at least some extent in order to reveal the text. Our first study showed that this style of presentation enhanced learning by 20%. Interestingly, we also found that learning of factual information (which was *not* illustrated by the illustration) was improved just as much as spatial information (which was illustrated by the illustration). This finding is

especially encouraging, because it means that the active style increases all learning, not just that directly related to the illustration.

The second study attempted to determine if either the act of moving the cursor onto the illustration or the fact that the text appeared closely adjacent to the illustration was the critical feature in enhancing learning in the integrated style. The study found no differences between these two conditions and the original style. These results, taken in conjunction with Study 1, suggest that the critical feature is simply that the user must interact with the illustration at some point in the study period. Thus, our findings support Wright, Hull and Black (1990), and extend their conclusions: learning is enhanced when the user either encounters an illustration unexpectedly, or is required to interact with the illustration.

The present studies support the following design guidelines concerning illustrations when the goal of the system is to inform the user:

- Have the user interact in some way with the illustration. Simply presenting a captioned illustration will not aid learning.
- Having the illustration appear unexpectedly will also enhance learning, but violates usability principles. Nonetheless, the technique could be effectively used in moderation.
- Reverse the usual order of saliency of text and illustrations. That is, make the illustration what users

see first, and have them perform an action to bring up the associated text.

- Even if some of the information is not directly related to an illustration, associate it in a meaningful way with other text that is directly related to the illustration. Learning for the associated information will be enhanced, as well.

Additional work needs to be done to extend the basic finding to other uses of illustrations. For example, it seems likely that the integrated interaction style might be more effective than paper media for teaching definitions, contrasts and comparisons. Consider a person who wants to learn the names of and differences among various styles of architecture. The learner might explore diagrams or pictures of various examples. As they move the cursor over the illustrations, textual information appears which points out the characteristics of each style. An additional challenge would be extending this work to interactive animation or videos. For example, research done by Pane, Corbett and John (1996) showed that students did not learn declarative information any better from multimedia educational software than they did from static text and images. Clearly, there is much that needs to be understood before we can take full advantage of all that multimedia can offer toward enhanced learning and information seeking.

5. REFERENCES

- Black, A., Wright, P., Black, D. and Norman, K. 1992, Consulting on-line dictionary information while reading. *Hypermedia*, 4 (3), 145-169
- Egan, D. E., Lesk, M. E., Ketchum, R. D., Lochbaum, C. C., Remde, J. R., Littman, M. and Landauer, T. K. 1991, Hypermedia for the electronic library? CORE sample results. *Hypermedia '91 Proceedings*, 299-312.
- Glenberg, A. M. and Kruley, P. 1992, Pictures and anaphora: Evidence for independent processes. *Memory & Cognition*, 20 (5), 461-471.
- Harrison, S. M. (1995) A comparison of still, animated or nonillustrated on-line help with written or spoken instructions in a graphical user interface. *Human Factors in Computing Systems*, CHI 95, 82-89.
- Instone, K., Brown, E., Leventhal, L. and Teasley, B. 1993, The challenge of effectively integrating illustrations into hypertext. *EWCHI '93 Proceedings*, Vol. 1, 78-86.
- Instone, K., Teasley, B. and Leventhal, L. M. 1993, Empirically-based re-design of a hypertext encyclopedia, *INTERCHI '93 Proceedings*, 500-506.
- Leventhal, L. M., Teasley, B. M., Instone, K., Rohlman, D. S. and Farhat, J. 1993, Sleuthing in HyperHolmes: Using hypertext vs. a book to answer questions. *Behaviour and Information Technology*, 12 (3), 149-164.
- Ogozalek, V. Z. 1994, A comparison of the use of text and multimedia interfaces to provide information to the elderly. *Human Factors in Computing Systems*, CHI 94, 65-71.
- Pane, J. F., Corbett, A. T. and John, B. E. Assessing dynamics in computer-based instruction. *Human Factors in Computing Systems*, CHI 96, 197-204.
- Teasley, B., Leventhal, L., Blumenthal, B., Instone, K. and Stone, D. 1994, Cultural diversity in user interface design: Are intuitions enough? *SIGCHI Bulletin*, 26 (1), 36-40.
- Waddill, P. J. and McDaniel, M. A. 1992, Pictorial enhancement of text memory: Limitations imposed by picture type and comprehension skill. *Memory & Cognition*, 20 (5), 472-482.
- Wright, P., Hull, A. and Black, D. 1990, Integrating diagrams and text. *The Technical Writing Teacher*, 17 (3), 244-254.