



JUS

JOURNAL OF USABILITY STUDIES

Vol. 3, Issue 4, August 2008, pp. 189-205

Unexpected Complexity in a Traditional Usability Study

Tharon W. Howard

Usability Testing Facility

Director

Clemson University

812 Strode Tower

Clemson, SC

USA

tharon@clemson.edu

Abstract:

This article is a case study of a demonstration project intended to prove the value of usability testing to a large textbook publishing house. In working with a new client, however, the research team discovered that what our client thought were simple problems for their users were actually complex problems that required the users to evaluate potential solutions in a surprisingly complex context of use. As Redish (2007) predicted, traditional ease of use measures were "not sufficient" indicators and failed to reveal the complex nature of the tasks. Users reported high levels of satisfaction with products being tested and believed they had successfully completed tasks which they judged as easy to complete when, in fact, they unknowingly suffered failure rates as high as 100%. The study recommends that usability specialists expand our definition of *traditional usability measures* so that measures include external assessment by content experts of the completeness and correctness of users' performance. The study also found that it is strategically indispensable for new clients to comprehend the upper end of complexity in their products because doing so creates a new space for product innovation. In this case, improving our clients' understanding of complexity enabled them to perceive and to take advantage of a new market niche that had been unrealized for decades.

Keywords

complex problems, visual vs. verbal design, complex tasks, working with new usability clients, methodological triangulation, usability measures, evaluating complex systems, writing handbooks, complex visuals



Copyright © 2007-2008, Usability Professionals' Association and the authors. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. URL: <http://www.usabilityprofessionals.org>.

Introduction:

In "Expanding Usability Testing to Evaluate Complex Systems," Redish (2007) observed, "Ease of use—what we typically focus on in usability testing—is critical but not sufficient for any product. Usefulness (utility) is as important as ease-of-use. If the product does not match the work that real people do in their real environment, it may be an easy-to-use solution to the wrong set of requirements" (p. 104). This study provides empirical evidence that validates Redish's observation. The traditional ease of use indicators in this study, both quantitative and qualitative, suggested that users found the prototype handbook my colleagues and I were testing to be an "easy-to-use solution" to their perceived needs. What's more, our client was entirely satisfied with the ease of use data we provided and didn't require that we use additional methodological triangulation in the study. As is often the case with clients new to usability studies, the clients were more concerned with "whether users liked using the product" than they were with what the study might be able to tell them about accuracy and functionality. This common attitude among clients who are new to usability studies can make it difficult for usability professionals to justify the additional expense and complexity of research designs that collect data on "utility" as well as ease of use (Redish 2007). However, had this study's design solely focused on what the client told us they wanted and had we only used traditional measures, we would have failed both our client and the users. Consequently, this case will hopefully serve as an argument that can be used to show clients why we need to go to the additional expense of methodological triangulation. What's more, this case is also a "success story" that can hopefully be used to illustrate how helping clients understand complexity they didn't know existed in the use of their products can create a space for innovation and creative thinking. In other words, this is a story about how discovering unexpected complexity enabled our client to release what has become a tremendously successful product into a market that had been dominated for decades by a traditional type of product design.

Definitions of complex problems vary, but according to Michael Albers (2003), "In complex problem solving, rather than simply completing a task, the user needs to be aware of the entire situational context in order to make good decisions" (p. 263). In doing a comparative usability test of two handbooks for students in freshman composition classes, this study provided a very interesting case in point of Albers' statement. Students thought the writing handbooks were easy to use, but they chose incorrect solutions to problems, because the handbooks did not help them understand the complexity of their task and did not provide guidance on how to choose wisely within that complexity.

This case study is also particularly interesting because we worked with textbook publishers. The computer industry has long benefited from usability studies, but the textbook industry has traditionally relied on reviews by experts. They rarely take time to gather user experience information from the actual users of their products, i.e., the students. In this project, my colleagues and I worked with the Director of Research and Development for a major textbook publisher, Allyn & Bacon/Longman, to design a study that sought to demonstrate that usability testing of textbooks with actual users could be used both (a) as a means of helping authors to produce more usable books and (b) as a means of helping acquisition editors and senior management evaluate whether or not to invest the resources needed to take the product to market.

Initially, the study had three goals:

- Convince textbook publishers of the value of usability testing.
- Compare a new visual type of handbook against a traditional version.
- Help find and fix problems in the new type of handbook.

However, as the study progressed and we learned more about our new client's understanding of the complexity of the tasks users were expected to perform with handbooks, a fourth goal emerged. We were able to demonstrate for the client that aggressively seeking to understand complexity in a product that they hadn't known was there simultaneously enabled them to see new market opportunities and helped create a space for creative thinking about a product that had been on the market for decades.

Visual versus Verbal Designs

One of the principal ways which Longman wished to distinguish its new handbook from the long-time market leader was to attempt to primarily deliver the content for the book visually rather than verbally. Our particular study compared the usability of a brand new grammar handbook designed entirely around the visual delivery of grammar and other writing conventions with a grammar handbook that has been the market leader and that is extremely verbal in its delivery of content.

As Figure 1 illustrates, Diana Hacker's *A Writer's Reference* (2006, p. 236) depends mainly on prose discussion to deliver the content it covers. Grammatical rules and conventions are stated in the traditional vocabulary of grammarians. Examples of the rules being discussed are provided, and these are usually set off from the prose discussion by maroon-colored bullets and boldfaced font. Occasionally, a maroon caret symbol or ^ is used beneath a line of text to call attention to an important punctuation. Alternatively, an italics font is used to call attention to key words in the example.

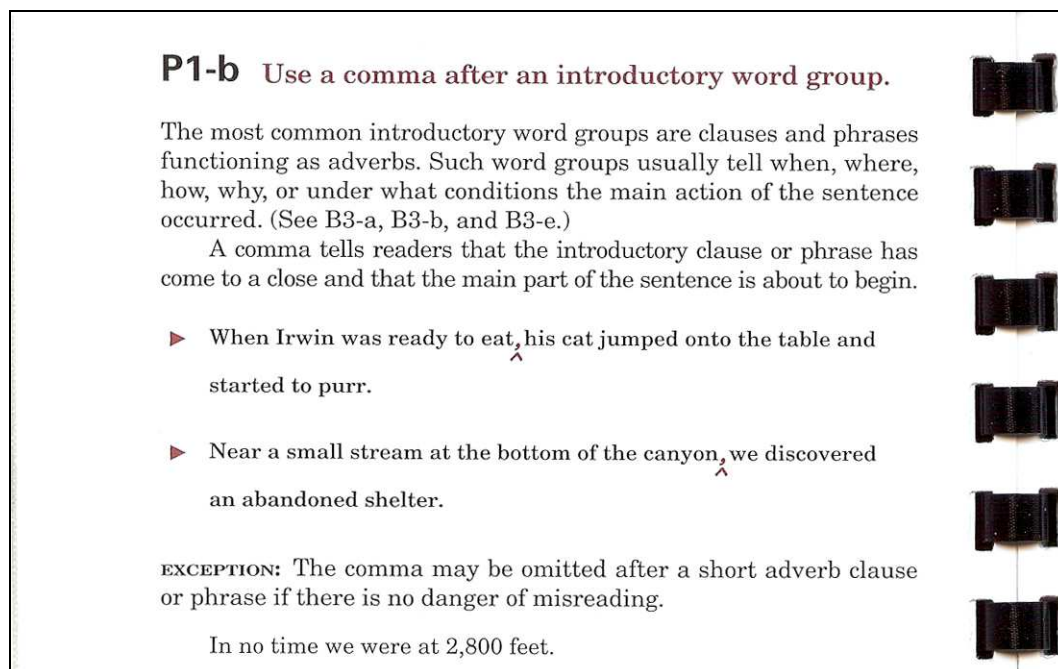


Figure 1. Sample page from market leading handbook

The Longman prototype handbook, however, seeks to minimize the amount of prose used to deliver content as well as grammatical terminology. As Figure 2 shows, the Longman prototype uses a variety of visual techniques to deliver content. Color coding is used to call attention to the differences in locations where phrases may be added in a basic sentence and then to show what those phrases look like and how they are punctuated. Also, rather than using terminology like "adding a medial modifier to an independent clause," the language here is much simpler and accessible to a non-specialist audience.

USING COMMAS TO ADD ADDITIONAL INFORMATION TO THE MAIN IDEA OF A SENTENCE

THE PATTERN

You can add additional information to a sentence at the beginning, middle, or end:

Most days, I make time to play outside. , usually in the garden or on a bike or taking a walk.

, together with my friends, , just a little bit of time, , to roam and jump and breathe,

You can add the suggested insertions (some of which are adjective phrases, some adverb phrases) to the big sentence above to build the following:

Most days, I make time to play outside.

Most days, I make time to play outside, usually on a bike or taking a walk.

I, together with my friends, make time to play outside.

I make time, just a little time, to play to play outside.

Most days, I make time to play, to roam and jump and breathe, outside.

NOTICE THE PATTERN OF COMMA USE AROUND THE INSERTED INFORMATION:

- When you add information at the beginning of a sentence, put a single comma after what you add.
- When you add information in the middle of a sentence, put a comma before and a comma after what you add.
- When you add information at the end of a sentence, put a comma before what you add.

432
PART 9 WRITTEN DELIVERY : PUNCTUATION

Figure 2. Sample page from prototype handbook

In terms of its approach to MLA documentation, the Hacker handbook approached content delivery in essentially the same fashion as was illustrated in Figure 1, using the same page layout and conventions to deliver its content. Because of its visual orientation, however, the Longman prototype used a different visual to illustrate to students and users how to prepare a works cited entry (see Figure 3).

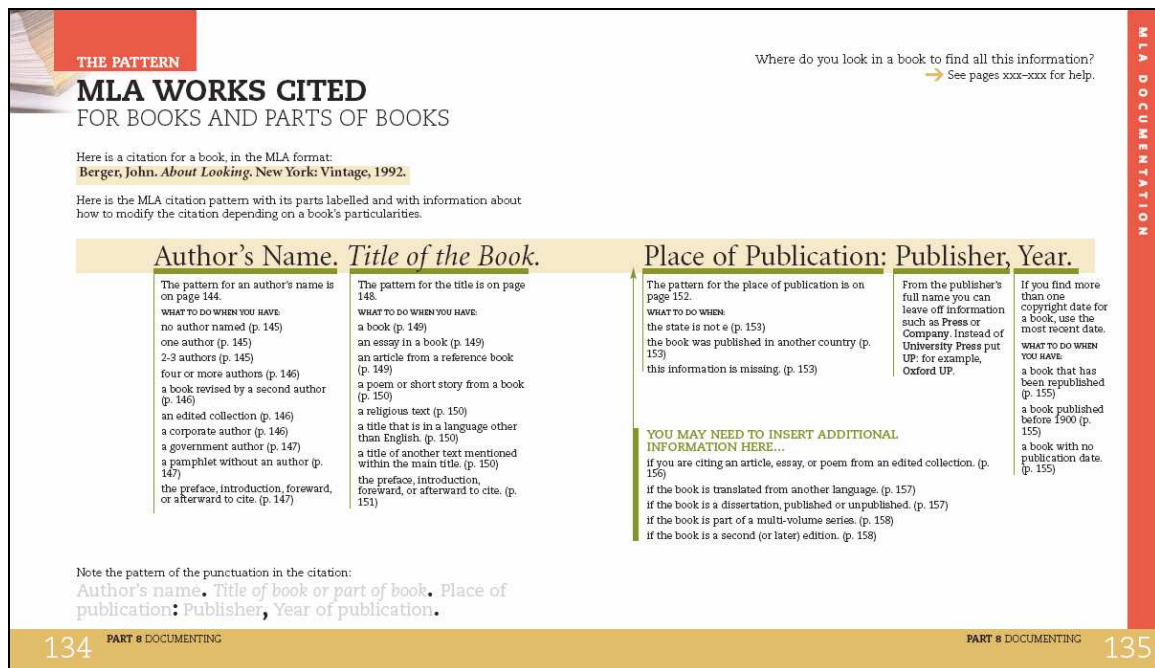


Figure 3. MLA works cited visual in prototype

Our study sought to determine whether the traditional verbal or the newer visual approach was more usable.

Methodology

The procedure used in this study was a traditional think-aloud protocol analysis. There were three major parts to the study:

1. Pre-test interview: where demographic and background data about users were collected.
2. Think-aloud protocols: where users were presented with scenarios requiring that they complete tasks and say aloud what they are thinking and doing as they perform the tasks.
3. Post-test interview: where users were asked to reflect critically on their experiences and to compare the texts examined.

Pilot testing was conducted on the instruments to ensure that the questions, instructions, and scenarios were understandable and to verify that the data collection instruments functioned properly, but data from the pilot testing are not used in this article.

Pre-test interview and subject profiles

After going through the informed consent statement agreement and giving their permission to be videotaped, participants in the study were asked to participate in a pre-scenario interview. This interview collected basic demographic information about participants' experience with high school English classrooms and helped us gauge whether or not the participants were representative of typical freshman composition students.

Table 1 provides a breakdown of the 12 participants who were recruited from 6 different composition classes and were paid \$75 for their participation. Because grammar handbooks are used in both 2-year and 4-year colleges, 6 of the participants (4 males and 2 females) were from a 2-year community college, and 6 of the participants (3 males and 3 females) were from a 4-year university. All of the participants were either 18 or 19, all 12 were in their first

semester of college, and most importantly, all 12 participants were currently enrolled in their first college-level composition course.

Table 1. Personal Background Information

User #	Age	Gender	Race	Major	High School GPA	4.0 Scale
1	18	M	A	Industrial Engineering	4.2/5.0	3.36
2	18	F	C	Communication	4.8/5.0	3.84
3	19	M	C	Undeclared	4.02/5.0	3.22
4	18	F	C	Spanish & International Trade	4.24/5.0	3.4
5	18	M	A	Computer Science	3.7/4.0	3.7
6	18	F	A	Chemistry	3.5/4	3.5
7	19	M	C	Univ. Transfer-Electrical Engineering	3.84/4.0	3.84
8	19	M	C	Univ. Transfer-Electrical Engineering	3.6/4.0	3.6
9	19	M	C	Undeclared	3.4/4.0	3.4
10	18	M	C	Univ. Transfer-Business Management	2.8/4.0	2.8
11	18	F	C	Univ. Transfer-Business Management	3.2/4.0	3.2
12	18	F	C	Univ. Transfer-General Studies (wants to go to Medical School)	3.2/4.0	3.2
TOTALS	18.3					3.42

A=African American

C=Caucasian

We also collected data about grades, SAT scores, majors, and other information in order to show that users were fairly representative of freshman composition users in the Southeast.

Scenarios and tasks for think-aloud protocols

After collecting this basic information about the participants, we introduced participants to the scenarios. Naturally, participants were instructed to talk out loud and to verbalize their thoughts as they attempted to use the textbook to perform the tasks provided. To help us track what they were observing on the pages, participants were instructed to point at the text as they moved through it and to read aloud when they were reading text. If users did not speak for more than 5 seconds, they were prompted by the test administrator and asked to explain what they were thinking. Also, *active intervention protocol* techniques were used to question users about their location in the text, what they were seeing, and how they felt about the material. Participants' comments were videotaped and coded, helping us to identify whether or not the page layout techniques, navigation systems, and other features of the text assisted or impeded users' ability to perform the tasks.

Our student users worked on three scenarios of increasing complexity:

1. Putting a complex source into correct MLA style (the *citing sources* scenario)
2. Identifying non-trivial comma errors (the *using punctuation* scenario)
3. Evaluating the acceptability of sources based on information about a specific assignment and the audience for the piece they would be writing (the *evaluating sources* scenario)

Each scenario increases the complexity of the task to be performed. The citing sources scenario essentially asked users to follow a model in order to complete a task. Identifying non-trivial comma errors was slightly more complex because it asked users to apply rules to a situation and to make a judgment. And the final task, evaluating possible sources for a library research paper, was the most complex because it required an understanding of the rhetorical situation in which the sources would be used. Users had to make a judgment about the appropriateness of

a source based on the exigency for the research paper, the audience for the piece, and wide variety of other environmental factors.

Participants used both handbooks for each scenario so they could compare the two handbooks. However, in order to control for first-use bias, the research team alternated which handbook they used first in each scenario. This ensured that both handbooks were used first an equal number of times.

Citing sources scenario

In the first scenario, users were asked to assume that they were working in their current composition class on a research paper. The scenario required that they create a works cited entry using MLA style. The researchers provided books with passages marked in them that users had quoted in the hypothetical research paper they had written.

During the pilot testing, we discovered that users would not actually use the handbooks thoroughly if the works cited entries were simple, single authored books. Because of their previous experiences writing research papers and creating works cited entries, it was necessary to challenge users with difficult and unusual citation tasks that actually required them to use the handbooks to find citation models that were unfamiliar to them. For part one of the citing sources scenario, the following text was used for the works cited entry:

Adobe Systems, Inc. *ADOBE PHOTOSHOP 5.0: Classroom in a Book*. San Jose, CA: Adobe Press, 1998.

For part two, users were instructed to use the other handbook, and the following text was used for the works cited entry:

Baecker, Ronald, Ian Small, and Richard Mander. "Bringing Icons to Life." *Human-Computer Interaction: Toward the Year 2000*. 2nd ed. Eds. Ronald Baecker, et al. San Francisco: Morgan Kaufman, 1995. 444-449.

At the end of each part, users were asked to rate the ease of use for the handbook using the following scale:

Very useful	Useful	Rarely useful	Not useful
-------------	--------	---------------	------------

Users were then asked to explain their rating.

Using punctuation scenario

In this scenario, users were asked to identify comma errors in a paragraph and to provide the page numbers from the handbooks that provided information about the correct comma usage. We presented users with a sample student essay pregnant with comma errors that were based on the 20 most common errors found in 3,000 college essays by Connors and Lunsford (1992; see also Smith, 2006). We identified four potential comma problems for the users and then asked them to locate information in the handbooks that told them:

1. if a comma was required at the location indicated,
2. if no comma was required, or
3. if the comma was optional.

This allowed us to collect data on how easy it was for users to locate specific comma usage information.

In part one, users were asked about comma usage in the places indicated by the four numbered circles in the following paragraph:

In America^① it is quite possible to live in a cocoon^② oblivious to the world around you. Confined living situations and close-knit social structures can prevent an individual from ever experiencing a reality,^③ outside of his or her own. Through an artistic medium such as photography one can get a glimpse of a world far removed. Gordon Parks photographer, artist^④ and writer, was a liaison between those Americans in one world and their fellow citizens who subsisted in a completely different one.

In part two, a similar paragraph was used for the second handbook. After completing each part of the scenario, users were asked to rate the ease of use for each handbook using the same ease of use scale as the citing sources scenario.

Evaluating sources scenario

The third and final scenario gave users a research paper prompt that included both a research topic and a specific audience for the paper. Users were then given possible sources for the research paper and asked to use the handbooks to indicate if the source was acceptable, unacceptable, or if more information was required. Users were again instructed to provide the page numbers from the handbook that enabled them to make their determinations. And once again, we collected ease of use data on this task. However, while the data from this third scenario were of interest to the client in terms of making recommendations on how to improve their textbook, the focus of this article is on complexity, and most of the complexity issues resulted from the citing sources and comma usage scenarios.

Controlling for print quality bias

Because the scenarios required that the students use and compare both handbooks and because one text was an unfinished draft, care was taken to ensure that texts used were of comparable finished quality. The materials used in the study were color copies of Longman's forthcoming *DK Handbook* and color copies of excerpts from Diana Hacker's *A Writer's Reference* (2006). Efforts were made to provide equivalent sections of Hacker's text so that neither text users received was complete. Both texts were cut to size, printed on facing pages, and plastic-comb bound. The texts were divided into the three sections, separated by Post-it[®] note tabs that were labeled *Commas*, *Eval. Sources*, and *MLA*. Throughout the study, the test administrator referred to **both** of the texts as "prototypes" under development and did not reveal to participants that the Hacker text had already been previously published. To further disguise the fact that one handbook was an unfinished draft, while the other represented excerpts from a published text, the test administrator labeled each copy with initials only. Throughout the study, the Hacker copy was referred to as *the HC handbook* or *the HC prototype*. The *DK Handbook* was only referred to as *the LP handbook* or *the LP prototype*.

Post-test interview

Having used both handbooks to complete tasks involving MLA documentation, comma usage, and evaluating sources, participants were asked a series of questions that allowed them to reflect critically on their experiences. Users were asked to compare the handbooks, to make recommendations for improving the handbooks, and to indicate which of the handbooks they would recommend to their teachers and why. Data obtained from the post-test interview as well as the scenarios are discussed below.

Findings

This section discusses the general findings and observations of the study.

Users preferred the visual approach

As Redish has suggested, traditional ease of use measures alone gave us no real sense of the complexity involved in the tasks. Overall, users reported that they preferred the DK prototype's visual ease of use to the more verbal approach used in Hacker. When asked to rank the "overall" ease of use for the two texts after they had actually used both handbooks, 9 of the 12 users preferred the DK prototype, and 9 of the 12 indicated that they would recommend it to their teachers for their entire class. Users recommending the DK prototype also appeared to have a stronger preference for their recommendations than the 3 recommending Hacker. Users were asked to indicate the strength of their preference on a scale from 1 to 10 where 1 indicated that they thought the text was "slightly better" and 10 indicated "vastly superior." The 9 users recommending DK averaged 7.44 (standard deviation was 2.35), and the 3 recommending Hacker averaged 6.00 (standard deviation was 1.0). See Table 2 for the range of scores users gave.

Table 2. Strength of Preference

	Strength of Choice	Strength of Choice
	DK	Hacker
	7	6
	5	5
	3	7
	10	
	9	
	9	
	9	
	9	
	6	
Avg.	7.44	6.00
Std. Dev.	2.35	1.00

During the post-test, users were also asked to give their overall ranking of the ease of use for both texts in terms of finding information (see Table 3). On a 4-point scale where 4 was "Easy" and 1 was "Difficult," the DK text received an average score of 3.33 (with a standard deviation of 0.78), and the Hacker excerpt received an average score of 2.33 (with a standard deviation of 0.89). And once again, the low standard deviation for the DK scores here are noteworthy since only 2 of the 12 users gave the DK text a score of 2 (or "somewhat difficult"), while the remaining users either gave the text a 3 ("somewhat easy") or a 4 ("easy"). Users' evaluations of the Hacker excerpts were slightly more varied, resulting in the larger standard deviation. However, the difference here is notable when one considers that only 5 users gave Hacker a score of either "easy" or "somewhat easy," and the remaining 7 users gave Hacker a score on the "difficult" side of the scale.

Table 3. Overall Ease of Use for Finding Info

	Overall	Overall
	DK	Hacker
	4	3
	4	2
	3	3
	2	3
	4	3
	3	2
	4	2
	4	2
	3	1
	2	4
	3	2
	4	1
Avg.	3.33	2.33
Std. Dev.	0.78	0.89

Users failed at tasks, but didn't realize it

Yet, while users' clear preference for the DK text and their overall "ease of use" scores are suggestive, it would be an error to conclude that the DK prototype's visual approach was more "usable" than a verbal approach. The ease of use evaluations above do not give a complete picture of the usability of the texts because the users' evaluations must also be considered in light of the question of whether or not users were actually able to complete the task "successfully" or not. In other words, users may initially give a positive evaluation of the ease of use for a product that they thought had helped them complete a task, but if a text misled them by allowing them to believe that they had finished the task when, in reality, the task was only partially completed, then the users' initial assessments are less valuable as a measure of usability. It is at this point that the issue of complex problem solving manifested itself in our study, and it is by means of content experts' assessment of the quality of the users' performance that researchers who are working with new clients can observe when complex problems may be disrupting the findings in a traditional usability study.

In this study, *both* the DK prototype and the Hacker excerpts failed the users when it came to successfully completing acceptable works cited entries for the works provided. All 12 users failed to provide a works cited entry that would have been judged satisfactory by college-level composition instructors. Even if one takes into account some of the information many instructors would consider optional—such as the use of "et al" for multiple authors or editors, the omission of words like "Press" and "Inc." from publishers' names, or the decision of whether or not to include the initials for the state after giving the city's name—even without these, users in this study omitted critical information necessary for a complete, acceptable citation. For example, users failed to list the authors of an article in an anthology, they failed to list a title of the essay, they failed to include the number of an edition, they failed to provide the page numbers for articles, and so on.

Users failed to recognize the complexity of their situations

Admittedly, the books that users were tasked with citing were challenging, but a strength of the study was that the task was also realistic. One of the works users had to cite was a corporate author where the corporation was also the publisher. The other was a book chapter with three authors published in the second edition of an anthology edited by four people. It was necessary to challenge the users so that they would actually *need* to use the handbooks to complete the tasks, and it is the case that all the information needed to cite both of the texts is provided in both handbooks. The books were intended to support precisely this sort of challenging citation, so if the handbooks were to be considered truly usable, it seems legitimate to have expected that users should have experienced more success than the total failure we observed.

Furthermore, we saw similar performance issues in the responses to the punctuation scenarios. Although the findings were less problematic from a performance perspective, our study found that users consistently failed to correctly indicate when the use of commas was required, not required, or optional, and they also failed to provide the correct page number from the texts where they obtained the information. For example, 11 of the 12 users incorrectly stated that a comma was required rather than optional after short, 2-word introductory clauses, and once again, this finding was observed for both handbooks. However, the users were completely unaware of this deficiency in their performance and did not consider this factor when they assessed the "usability" of the handbooks. It was the discovery of these extremely poor performance indicators *for both texts* that initially led us to question what might have been at issue. Had only one text failed, then this might have suggested that it was the delivery technique used by that product that was at issue. However, the failure of both products was the first real clue that complex problems were at issue.

Why did users fail?

Unfortunately there is no single, obvious, one-size-fits-all explanation that describes why some users struggled with the comma sections in the handbooks or that can adequately illustrate why all 12 users failed to produce appropriate works cited entries. Several factors contributed to users' problems:

- Users scanned pages for examples that matched the mental models they had for patterns and only stopped to read material when they found patterns that matched those models.

- They thought the problem was simple and didn't look beyond the first solution—even when it wasn't enough.
- They relied on bold headings and skipped the paragraphs.
- The visual manual tried to combine too much information in one graphic.
- The authors of the manuals didn't understand their users' mental models.

To illustrate the difficulties here, it may be worthwhile to examine the ways users attempted to address the question of whether or not a comma is required or optional after the phrase “In America” in the following sentence:

In America it is quite possible to live in a cocoon.

The correct answer to this question is that the comma is optional, which is explained on page 433 in the DK prototype and on page 236 in Hacker. However, only one DK user correctly gave “optional” as a response, and this user incorrectly identified the pages where the information could be found. All the other DK users incorrectly stated that a comma was required, but only 2 of those 11 gave page 432 as the page that indicated that a comma was required (see Figure 2 in the Background section). The other DK users gave page 428 as the page that contained the information because page 428 had examples of sentences that looked like the pattern. By contrast, all of the Hacker users correctly identified page 236 as the page with the information they needed, but only because the only examples available were on page 236.

Users were satisfied by the first simple solution

A closer examination of the process users followed may help to account for some of this difference, but the principal point to be made here is, users assumed that they were dealing with a simple problem, so once the handbooks suggested to them that they had found the simple solution, they didn't look any further for more complex answers. Yet, deciding whether or not a comma is required after the introductory clause “In America” required that users needed “to be aware of the entire situational context in order to make good decisions” (Albers, p. 263). Whether or not the comma is required depends on the rhetorical situation in which the sentence occurs (it is required in a formal piece of discourse such as a business proposal and may be omitted in an informal medium such as a letter to a friend). However, because the handbooks functioned acontextually, they never signaled to users that context might be a factor in their decision-making processes.

For example, the DK users would read the sentence in question and identify the introductory clause pattern (a few users actually called it an “intro clause”). They would then quickly decide that they were dealing with a question about commas and they were very successful at identifying where they should go in the text to locate more information. Because printed pages were used in this study, eye-tracking systems that could have confirmed where users were looking on the page were not available to the researchers. Nevertheless, all of the users appeared to read the large bold heads that said, for example, “Use commas to make numbers, place names, and dates clear.” Often these headers were all that users needed to decide whether or not they would find information they were seeking. In the main, users skipped prose passages. Once users found headers that suggested they might be close to the type of comma use they were seeking, they only scanned the examples on the page. They looked at examples in a very specific way—i.e., to decide if they could match the syntactic pattern they were seeking to the examples. Once they found the example that matched the pattern, the use of commas provided by the example was the only answer they felt they needed, and they rarely read any prose text to confirm the accuracy of their decisions. Unless there were additional examples or some other visual clues to suggest that the decision might be more complex, users assumed it was simple and read no further.

Users scan headers and examples and skip text

Although this scanning behavior was new to our clients, it will come as no surprise to usability professionals who have observed users of computer or software user guides. Given that users read headers and scanned for examples in order to decide whether or not to read a page or a section, it might be easy to assume that the visual approach used by the DK prototype would be far easier for users to scan (see Figure 2). However, it should be observed that the logic of Hacker's pages also enabled users to aggressively seek out examples on a page and then slavishly follow those examples. Hacker's pages use a maroon colored header to state the

comma usage (e.g., “Use a comma after an introductory word group”), and the pages set off examples from the rest of the text with a maroon bullet, a different font, indentation, and double-spacing. Users complained, sometimes vociferously, about how they “hated” Hacker’s small fonts and the fact that the page design made them flip around and read too much. But they were also careful to qualify this by saying that they were, ultimately, able to find the information they wanted. However, their performance on the question of whether introductory commas were required or optional was essentially the same as the DK prototype. This is startling because, even though Hacker states on the same page that there are exceptions when an introductory comma may be omitted (see Figure 1), these users did not read the exception to the rule. They read the header, they looked at the example, and they decided that they had all the information they needed in order to make a judgment.

The visual handbook tried to combine too much in one graphic

Another observation that appeared to have contributed to the problems users had with inappropriate citations and with users’ problems with correct comma usage had to do with the overuse of a single graphic or visual. It may be that the authors of the DK manuscript were limited to a strict page count, or they may have been restricted in the number of graphics they could use. However, the authors appeared to be attempting to force as much information as possible into graphics and visual elements like those illustrated in Figures 2 and 3. This led to visuals

- that actually produced errors,
- that confused users so they ignored the information, or
- that led users to complain that the book was “tangled up” and visually “messy.”

For example, user 11 complained that the visual shown in Figure 2 “threw” her because it was trying to do too much, and she went on to explain that the information around the base sentence “I make time to play outside” is trying to illustrate five different syntactic structures in the same visual. Similarly, user 10 complained that the “pattern pages were distracting” because the “pattern” pages were those that used visuals like those shown in Figure 2 and 3.

Because of its complexity, the graphic in Figure 2 appeared to contribute significantly to the performance errors and decisions not to read the text described in the previous section. In fact, it should be observed that few users actually looked at or commented on the graphic during the actual think-aloud protocol. The test administrator often had to take users to the pages in the post-test interview and ask them about the visual. During these reflective moments, users explained that they did, actually, glance at the visual; however, they complained that they decided that the graphic on 432 was “too busy” and would require too much effort to understand. Consequently, they skipped it. This was an unfortunate decision since six of the eight comma uses from the punctuation scenario needed information described on page 432.

It should be noted that this was not the case for the visual shown in Figure 4. Users observed the pattern, understood it, and commented favorably on its clarity. This led us to question why Figure 2 was ignored and Figure 4 was success.

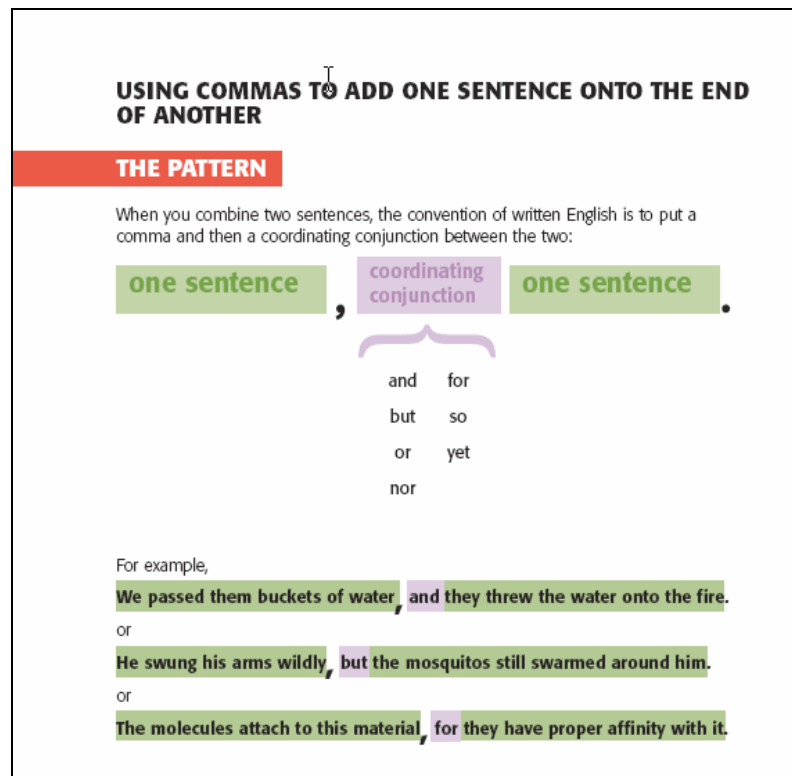


Figure 4. Successful single-purpose visual

Part of the answer we believe lies in the scanning behavior described in the previous section. Users scanned almost exclusively to match syntactic patterns in the examples. In Figure 2, the most prominent feature of the visual is the base clause or main idea, "I make time to play outside." The clauses or additional information that are added to the main idea are all visually subordinated to the large, uncolored, and uncluttered main idea. Yet, it is precisely this subordinated material which users needed in order to match the syntactic structures they were seeking. Additionally, instead of having to examine one structure and then decide if it matched the pattern they were seeking, Figure 2 required that they consider at least five patterns in the same visual. Conversely, the more successful Figure 4 does not subordinate elements, and it does not attempt to combine multiple patterns into one visual.

The visual as shown in Figure 3 was even more problematic than Figure 2, however. This is ironic because users were nearly unanimous in their positive comments about Figure 3 and how effectively they thought that the visual would help them to produce a properly formatted MLA works cited entry. However, this visual actually produced errors because it failed entirely to alert users to the problems faced when citing an article from a book, an edition, a corporate author, an edited collection, etc. Users were overwhelmed by the large font, yellow highlighting, and underlining of the pattern provided and failed to recognize that additional information would be required. Although once again the information they needed was provided, the treatment of the material in the visual misled users into thinking that they had adequately responded to the task. Users believed that, if they supplied the "Author's Name," "Title of the Book," "Place of Publication," "Publisher," and "Year," then they had provided all the data needed for the works cited entry, when, in fact, they had not. As a result, users commented positively on the visual; yet, every single user failed the task. It would be unfair to assert that the visual in Figure 3 was entirely responsible for this failure, and the next section discusses other factors that appeared to lead users of both the Hacker and DK texts to produce incomplete works cited entries. However, asking the visual in Figure 3 to accomplish too much certainly appeared to contribute to the problem, and once again, we observe that the failure of the information product to signal

to users that they were dealing with a complex problem resulted in positive user evaluations and poor user performance.

Users' task environments can be improved with scenarios

The failure of both the Hacker excerpt and DK prototype on the works cited task, and users' difficulties with commas, can all be partially attributed to incomplete understandings of users' goals and task environment. It is obvious that improving an author's understanding of the users' needs is likely to result in more usable information products, hence the mantra "Know Thy User." This is well understood.

What is often less well-understood is the role that authors play in actually constructing the use or task environment for users. All too frequently, we tend to think in terms of accommodation of users' needs, and we tend to overlook the important role that authors and designers play in the construction of users' task environment. Indeed, the movement away from theories of "user-centered design" in the 1990's toward "user-experience design" is largely a recognition of this complex negotiation between accommodation of users on the one hand and creation of user-experiences on the other. Successful texts and information products create roles and provide interpretive frameworks that users can deploy in order to successfully complete tasks and achieve their goals.

The evaluating source section of the DK prototype did this fairly successfully, and it did so mainly by using "If, Then" scenarios that users could play in order to understand some of the criteria that need to be taken into consideration when deciding whether or not a source is credible and relevant (note: see Flowers, Hayes, and Swarts's 1983 article "The Scenario Principle" for more details about this approach). Rather than attempting to provide linear, step-by-step procedures for judging the value of a source, the DK prototype used mini-stories to exemplify how the complex decision of whether a source is relevant can be made. In one of these "stories," students named Pedro and Aaliyah are described. Pedro is writing a research paper for his class about social actions the U.S. ought to take to prevent pandemics. Aaliyah is researching viral marketing techniques for a marketing company. Pedro and Aaliyah are both asked to judge whether an online news story about how viruses spread electronically. The handbook guides users through the characters' decision making processes to show why the same text is not appropriate for Pedro's situational context on the one hand, and completely relevant to Aaliyah's on the other. This section creates a clear role for users to play as they use the text and it also acknowledges the importance of context on decision-making, which the other sections of the handbooks did not adequately address or address at all. And because the users understand the role they are supposed to play, they also understand the logic they're supposed to follow in order to make the decisions necessary to complete the task at hand.

This did not pertain, however, for either handbook in the MLA sections. Both texts failed because both asked users to play roles they could not adopt. Both texts assumed, erroneously, that users begin the citation process with a clear understanding of the types of works being cited and thus had a simple problem of choosing the format that matched the type of work. However, the users in this study did not understand the context and so didn't differentiate between types of works. Generally speaking, the users in this study didn't know, for example, what a *corporate author* was, and two of the users decided that a collection of readings was "a reference book." If you, as a user, don't know what a corporate author is, you're not going to be able to use the information on Figure 3 of the DK prototype or the tables on pages 341 and 349 in Hacker's handbook.

When users don't know what decisions they need to make (e.g., how do I decide if I have a corporate author, government author, no author, etc.), then constructing a task environment like the one created in the evaluating sources section is a positive approach. The presentation of the information in terms of "agencies" with whom the students and users could identify helps users understand where to begin the decision-making process. The authors should also ask questions as headers, such as "How do I quote or paraphrase in my text?" or "How do I format an entry for a works cited, reference list, or bibliography?" This would create the same "context of use" found in the evaluating sources section, and it allows the authors to build a decision matrix or some other tree-type structure that could be used by students and teachers to make decisions about what type of text they have so that they can make an informed judgment about the appropriate works cited pattern to use.

Conclusion

This study provides empirical evidence that validates Redish's observation that "Ease of use—what we typically focus on in usability testing—is critical but not sufficient for any product" (2007, p. 104). The ease of use indicators in this study, both quantitative and qualitative, suggested that our users found the prototype handbook and its visuals to be an "easy-to-use solution" to their perceived needs. And our client would very likely have been entirely satisfied with the ease of use data we provided. However, had the study's design solely focused on what the client told us they wanted, we would have failed both our client and the users. The true complexity of the tasks would not have been revealed, the lack of "utility" in the handbooks would not have manifested itself (Redish 2007), and our clients would never have seen the need to "think outside the box" of traditional handbook design. They would have lacked the cognitive dissonance that motivated at least some of the creative thinking that ultimately led to the successful handbook they ultimately published.

Fortunately, because we were working with clients who trusted us not to run up the cost of our study needlessly, we introduced some additional elements to the research methodology used. As a result of these additional elements, we observed the following points:

- It is important not to be swayed by clients and to ignore methodological triangulation. The decision to track the quality and correctness of the users' performance was fortunate. Had it not been for the observation that users failed to have acceptable works cited entries, we would not have suspected that the problems were more complex than we had originally considered.
- Testing the market-leader and the prototype with the same scenarios helped the researchers to distinguish problems with information delivery techniques from confusion over the complexity of the tasks.
- Maintaining a singular purpose in visuals is important for success, even though complex information needs to be delivered. Use several visuals to convey different purposes.
- Users need to be made aware of the complexity of a problem before any potential solutions are introduced so they can look for additional contextual factors and use a recursive decision-making process.
- Modeling complex problem solving behaviors through the use of scenarios may lead to more effective performances from users when it's simply not possible to capture or replicate all of the potential variables in a situational context (Flowers, Hayes, and Swarts, 1983).

As the field of usability studies grows and usability professionals are required to work with and adapt to the needs of new clients in new fields, the likelihood of encountering complex problems masquerading as simple ones will increase. Our research methods need to accommodate this and ensure that the studies we design for new clients include content experts' assessment of the quality of the users' performance and, when possible, head-to-head comparisons with competing products.

Practitioner's Take Away

- With new clients usability specialists need to accommodate the likelihood of encountering complex problems masquerading as simple ones and ensure that the studies we design use sound methodological triangulation techniques, including content experts' assessment of the quality of the users' performance and, when possible, head-to-head comparisons with competing products.
- Users in this study assumed that they were dealing with a simple problem, and once they found what they thought was the simple solution, they didn't look any further for more complex answers.
- Users gave a positive evaluation of the ease of use for products that they believed had helped them complete a task when, in fact, the information products misled them, allowing them to believe that they had finished tasks that were only partially completed.

- To avoid giving users a false sense of success that can ultimately lead to poor user performance, visuals must signal to users that they are dealing with a complex problem.
- Maintaining a singular purpose in visuals is important for success, even though complex information needs to be delivered. Use several visuals to convey different purposes.
- Users need tasks that help them understand where to begin the decision-making process. When users understand the role they are supposed to play, they also understand the logic they're supposed to follow in order to make the decisions necessary to complete the tasks at hand.
- Modeling complex problem solving behaviors through the use of scenarios may lead to more effective performance from users when it's simply not possible to capture or replicate all of the potential variables in a situational context (Flowers, Hayes, and Swarts 1983).
- Helping new clients understand the upper end of complexity in their products can make it possible for designers to perceive new opportunities for product innovation and can help them create new strategies for market success. In this case, helping the client see that their competitor had failed to recognize the complexity involved in citing MLA sources and in responding to commas also helped them find innovative ways of entering a market that had been dominated by one product design approach for decades.

Acknowledgements

The author would like to thank Wendy Howard who collaborated on the data collection and analysis portions of this study. Thanks also belong to Alicia Hatter, Ginny Redish, and Barbara Mirel for the insightful feedback that was critical in producing the final version of this article.

References

- Albers, M.J. (2003). Complex problem solving and content analysis. In M.J. Albers and B. Mazur (Eds.) *Content and Complexity: Information Design in Technical Communication* (pp. 263-284) Mahwah, NJ: Lawrence Erlbaum Associates.
- Beyer, H. & Holtzblatt, K. (1998). *Contextual Design: Defining Customer-Centered Systems*. San Francisco: Morgan Kaufmann.
- Connors, R.J., & Lunsford, A. (1992). Frequency of formal errors in current college writing, or Ma and Pa Kettle do research. In R.J. Connors & C. Glenn (Eds.) *The St. Martin's Guide to Teaching Writing* (p. 398) NY: St. Martins.
- Flower, L., Hayes, J.R., & Swarts, H. (1983). Revising functional documents: The scenario principle. In P. Anderson, R.J. Brockman, and C. Miller (Eds.) *New Essays in Technical and Scientific Communication*, Farmingdale, NY: Baywood, 41-58.
- Hacker, D. (2006) *A Writer's Reference, 6th Edition*. NY: Bedford/St. Martin's.
- Jensen, B. (2007) Lead, wallow, or get out of the way, *User Experience* 6 (2), 28.
- Mirel, B. (2003) Dynamic usability: Designing usefulness into systems for complex tasks. In M.J. Albers and B. Mazur (Eds.), *Content and Complexity: Information Design in Technical Communication*, Mahwah, NJ: Lawrence Erlbaum Associates, 233-262.
- Redish, J. (2007) Expanding usability testing to evaluate complex systems, *Journal of Usability Studies* 2 (3), May, 102-111.
- Smith, B. (2006) Gordon Parks: Using photographs to spark social change. In J. Ruskiewicz, D. Anderson, and C. Friend (Eds.), *Beyond Words: Reading and Writing in a Visual Age*, NY: Longman-Pearson, 149-153.
- Vincente, K. & Rasmussen, J. (1992). Ecological interface design: Theoretical foundations, *IEEE Transactions on Systems, Man and Cybernetics* 22, 589-606.
- Wysocki, A.F., & Lynch, D. (2008). *The DK Handbook*. NY: Longman.

About the Author

**Tharon W. Howard**

Dr. Howard has directed Clemson University's Usability Testing Facility since 1993 and teaches graduate seminars in user experience design, visual communication, and digital rhetoric(s). He is the listowner for UTEST and has received awards from the UPA and the STC recognizing his contributions to the field of usability studies.