

---

## What Can Document Designers Learn from Usability Testing?<sup>1</sup>

Karen A. Schriver  
Carnegie Mellon University

*Although methods for evaluating the quality of documents have existed since the 1930s, the document design community has attended most seriously to the theory and practice of usability testing in the past decade. One aspect of usability testing that has not yet been well explored lies in what it can teach writers about readers. This paper presents a study that evaluated a method for improving writers "to anticipate readers'" needs. The method, called "reader-protocol teaching," was developed from readers' responses that were collected during usability testing. The study shows that writers taught with the reader-protocol teaching method improved significantly in their ability to take the reader's point of view when planning to revise. Writers taught with the reader-protocol method significantly increased in their ability to diagnose readers' problems caused by textual omissions, characterize problems from the reader's perspective, and attend to global-text problems. These findings show that extensive practice in analyzing readers' responses to texts can have important cognitive benefits for writers.*

ALTHOUGH RESEARCHERS AND PRACTITIONERS from around the world share the goal of producing quality documents, there have been almost no methods developed for improving writers' abilities to design documents that work for readers. My purpose here is to describe briefly a method I developed for teaching writers to anticipate readers' needs for functional texts (for a more extensive discussion, see Schriver, 1992). The idea for this teaching method grew out of research I conducted at Carnegie Mellon University's

---

1 Appeared in *Studies of Functional Text Quality*. 1992. Edited by Henk Pander Maat and Michael Steehouder (pp. 141-157). Amsterdam: Rodopi Publishers. (Also available from Rodopi in Atlanta, Georgia.) Reprinted by permission of the author.

Communications Design Center (CDC)—a non-profit organization dedicated to basic and applied research in document design. During the 1980s, researchers at the CDC developed and elaborated a method called protocol-aided revision (Swaney, Janik, Bond & Hayes, 1981/1991; Schriver, 1984, 1991). Its purpose was to enable writers to employ direct feedback from readers to guide revision activity. Through asking readers to think-aloud as they worked with a document and a machine, writers are able to capture readers' real-time cognitive processing of the text and identify the aspects of functional texts that create difficulties for readers.

In the mid 1980s, there was a core of writers at the CDC who had taken part in much of the document design and usability testing research at Carnegie Mellon and who for a number of years had been using protocol-aided revision to evaluate functional texts. I observed that these experienced writers and usability testers seemed to be much better at planning text than were writers who had years of on-the-job editing experience. I wondered why these writers were so good and thought that perhaps it had to do with their extensive involvement with watching readers-in-action, experience that usability testing provided them. My observation led me to believe that perhaps it was the repeated exposure to how readers actually respond to text that changed the way these writers considered the audience during planning. I speculated that writers using readers' feedback to revise may have acquired a sensitivity to audiences' needs that writers experienced only in editing finished products could not acquire.

To explore my intuition that perhaps repeated experience with readers' feedback to revise might generalize to situations in which no reader feedback is available, I designed a teaching method to provide writers with practice in analyzing readers' responses to poorly written functional texts, practice that was modeled on the experiences of writers using reading protocols to revise. The method, called reader-protocol teaching, employs readers' responses to illustrate what people do when they read functional texts, particularly when they fail to comprehend the writer's intended message.

In particular, I explored the hypothesis that extensive experience in interpreting readers' feedback (provided through transcripts of think-aloud reading protocols) would help writers to become more aware of how readers construct text. Several related questions motivated my inquiry:

- Would the reader-protocol method help writers notice and characterize readers' responses to text?

- Would the reader-protocol method work as well or better than more familiar methods such as audience-analysis heuristics, peer-response methods, or role playing?
- Would the sensibilities that writers acquire through the reader-protocol method transfer to new texts and to new genres?

## Method

### *Participants*

The participants were 117 college juniors and seniors from ten classes in "Writing in the Professions." Five classes served as an experimental and five as a control group. Students were enrolled in a variety of degree programs in humanities, engineering, or business management. The course was elective, and class size ranged from 12 to 22 students. Data reported here were collected from 117 students, 43 students in the experimental classes and 74 students in the control classes.

### *Design*

A pretest was given to both experimental and control classes early in the semester. During the semester, students were taught to anticipate readers' needs through either the reader-protocol teaching method or through a combination of methods, including audience-analysis heuristics, collaborative peer-response groups, and role-playing activities. Teaching for both experimental and control classes took place over about six weeks. Both groups were posttested about three-quarters of the way into the semester.

The study required that I develop *teaching*, *testing*, and *validation* materials, each of which is described below.

### *Teaching Materials*

The materials for the reader-protocol method consisted of ten lessons, each containing two parts:

1. a "*problematic*" text. That is, a poorly written text that will cause comprehension difficulties for the intended audience. I selected ten problematic texts composed by students; each text was one to four pages in length. The texts were elementary

lessons in operating a university computing system and were intended for freshmen, secretaries, and university staff. An important feature of these texts was that they did *not* contain spelling or grammatical errors. Instead, they had incomplete forecast or preview statements, poor definitions, unclear procedures, missing examples, misleading headings, ambiguous goal statements, weak summaries, and other “beyond the word-level or phrase-level problems.” Students who created the texts (senior English majors) had been asked to write to an audience of college freshmen who had never used a computer before. Students who created the texts were not enrolled in the classes in the study. Each text introduced a new topic (e.g., sending mail, formatting a report, creating a table, using an on-line card catalog for the campus library) and was written to inform and instruct a lay reader.

2. *a think-aloud reading protocol of a person trying to understand the text.* For each of the ten texts, a reading protocol was collected from a different member of the actual audience—

#### Commands for English Text

EMACS enables you to manipulate words, sentences, or paragraphs of text. In addition, there are commands to fill text, and convert case.

Editing files of text in a human language ought to be done using Textmode rather than Fundamental mode. Invoke M-X Text Mode to enter Text mode. See section 20.1 [Major Modes], p. 85. M-X Text Mode causes Tab to run the function ^R Tab to Tab Stop, which allows you to set any tab stops with M-X Edit Tab Stops (See section 11.3 [Indentation], p. 46). Features concerned with comments in programs are turned off except when explicitly invoked. Automatic display of parenthesis matching is turned off, which is what most people want. Finally, the syntax table is changed so that periods are not considered part of a word, while apostrophes, backspaces and underlines are.

If you are editing input for the text justifier TEX, you might want to use TEX mode instead of Text mode. See the file INFO: ETEX. INFO.

**Figure 1.** An example from one of the ten reader-protocol lessons.

freshmen and secretaries learning to use a computer. The protocols revealed a wide variety of understanding and usability problems—problems that poorly written instructions often create for readers. The ten protocols were typed so that students could easily distinguish readers' responses from the original text.

The ten lessons were culled from a set of more than twenty texts on which think-aloud reading protocols had been collected. **Figure 1** presents an excerpt from one of the original problematic texts and a think-aloud reading protocol collected on that text. I chose the lessons based on how well they highlighted the ways that instructional texts can create problems for readers and make comprehension a painful process (Schriver, 1984).

### *Pretest and Posttest Materials*

Materials for the pretest and posttest were six naturally occurring expository science texts, each of approximately one-half page in length. The six texts were excerpts from the "Science" and "Medicine" sections of *Time* and *Newsweek* magazines; thus, they were intended for a general U.S. high school audience with an average reading level of grade eleven. These elementary scientific texts were not altered in any way. (They were, in fact, a subset of the texts used in the Thibadeau, Carpenter, and Just [1982] eye movement studies of the reading process.) The topics covered in the texts were: artificial hearts, babies' smiles, holography, flywheels, vitreous humor, and glial cells. Like the texts used in the ten lessons, the testing materials did not contain grammatical, spelling, or mechanical errors. Rather, they caused comprehension problems for readers because the writers failed to provide necessary contextual information, left out examples, presented confusing metaphors or analogies, provided illogical or faulty transitions, or packed information too densely.

### *Validation Materials*

To evaluate the accuracy of experimental and control writers' predictions of readers' responses during the pretest and the posttest, I identified the problems the testing materials created for a target audience of lay readers. To do so, I collected reading protocols from 30 freshmen trying to understand each of the six texts used in the pretest and posttest. The order in which the 30

freshmen read the six texts was counterbalanced. Two raters, who were graduate students in English, independently evaluated each of the 180 protocols provided by the freshmen and analyzed the problems freshmen experienced in trying to understand the passages.

The raters evaluated the protocols for “referable” and “non-referable” problems. I defined referable problems as those that were triggered by an identifiable locus in the text, for example, “this section here makes the idea too hard to understand.” Non-referable problems did not have such a locus; rather, they appeared to reflect cumulative effects of many parts of the text,

*Ok, now I'm going to try. . . Commands for English Text.*

*EMACS enables you to manipulate words, sentences, or paragraphs of text. These commands sound like ones I'd use all the time—good. In addition, there are commands to fill text, and convert case. I don't know what it means to fill text, I guess it means putting data from one text into another. . . that is, filling the text with what you want in it. Well, I guess I'll soon find out.*

*Editing files of text in a human language human language? Boy that sounds strange, what could they be distinguishing here? Maybe computer language or machine language from human language? ought to be done using Textmode rather than Fundamental mode. Well, I don't know what textmode or fundamental mode is, so how will I know which I'm in? Let's see... Invoke M-X Text Mode to enter Text mode. I won't do that because I do not have time to see the other section. That's terrible to tell me to . . . See section 20.1 [Major Modes], p. 85. M-X Text Mode causes Tab to run the function ^R Tab to Tab Stop, Now wait a minute, what is all this stuff? Sounds important, but I have no idea what it could mean. I know what tabs on a typewriter are, and I see a tab key here, but what the heck does running the function ^R Tab to Tab stop mean? Oh God, now they want me to look up another section . . . which allows you to set any tab stops with M-X Edit Tab Stops (See section 11.3 [Indentation], p. 46). . . Just to get started in this one. No I think I'll skip that. I'll deal with it when I have to use it. I thought this section was on how to edit: what does setting tab stops have to do with editing? Features concerned with comments in programs are turned off except when explicitly invoked. Hmm. . . a feature? Wonder what a comment feature is? How do they get turned off? Do I have to turn them off? Well I guess I can't turn them off if I don't know what they are. Automatic display of parenthesis matching is turned off, which is what most people want. What does that mean? How do I know if I'm like most people and want them turned off? I wish they'd tell me why I need to know all these hacker terms! Finally, the syntax table is changed so that periods are not considered part of a word, while apostrophes, backspaces and underlines are. Syntax table. . . no idea. Of course, I do know what the syntax of sentences are, so maybe it means something to do with certain commands EMACS considers acceptable. . . maybe certain actions in a row produce a correct syntax.*

*If you are editing input for the text justifier TEX, you might want to use TEX mode instead of Text mode. See the file INFO: ETEX. INFO. Well, I don't want to edit input for the text justifier mode, I just want to figure out how to manipulate words and paragraphs—like it says up here (points to the top of section). This stuff is too complicated and it's aggravating to read. . .*

**Figure 1 (cont).** An excerpt from a thinking-aloud reading protocol.

for example, “this whole thing is confusing.” Interrater reliability in judging referability was .914 (using Cohen’s Kappa). I used only the referable readers’ problems as a measure against which to judge experimental and control writers’ predictions of reader problems. Non-referable problems accounted for less than 10% of the problems experienced. I established the criterion that any referable problem that was experienced by three or more of the freshmen would be designated as a problematic text unit. Thus, if 10% of the 30 freshmen had a referable problem at the same text unit, I considered it enough of a problem to warrant a writer’s attention.

Three types of referable problems emerged: local (problems at the word-level), sentence-level (problems within sentences) and global (problems beyond the sentence, for example, “this text needs an example here”). Frequently, the same text area would cause problems of more than one type—problems that ranged from vocabulary difficulties to those of integrating new information within the context of given information. The goal in coding for referability was to provide a rough index of the frequency and location of reader problems and not to make theoretical claims about the nature of problems in text.

I used the location information about the number of readers who experienced referable problems to create answer templates for scoring writers’ predictions. In coding the protocols, a problem was defined as any statement that signaled confusion or misunderstanding of the text. The freshmen readers made statements like the following:

I do not understand how a baby could have an “inward growing grin.” Why would a psychiatrist study such a thing as grins? [The freshman was reading the text, *Babies’ Smiles*, which contained a discussion that described “sleepy smiles” as those in response to “external stimuli” and then distinguished the “sleepy smiles” from those referred to as “inward growing grins.”]

A ‘miniature nuclear furnace!’ Why would you use nuclear power inside a human being? Is this one of those metaphors or something? [The freshman, who was reading the text, *Artificial Heart*, was confused by a statement that said artificial hearts use a “miniature nuclear furnace” to keep them going.]

Using the readers' problems identified through analyzing the protocols, I created scoring templates for assessing writers' predictions of readers' difficulties. The protocols provided a more reliable measure of readers' difficulties than my intuitions about the problems the testing materials might have created for a lay audience.

## **Procedures**

### *Teaching Procedure for Experimental Classes*

On the first day, students in the experimental classes were provided one-half of the first lesson. That is, students were presented with one of the ten problematic texts. Students were told that the intended audience consisted of freshmen in college who had not used computers before, that is, readers unfamiliar with the topic of the text. Students were prompted to decide how well the text met the needs of the audience by employing the following procedure.

First, students read the text and underlined or bracketed words, phrases, sentences, sections—any size text unit that they felt would cause a lay reader trouble in understanding the text. To discourage students from underlining the whole text, students were asked to try to be certain that any text area they underlined or bracketed would cause only one problem for a reader. Then students were asked to assess each part of the text they marked and to diagnose what they thought the reader's problem would be. Students diagnosed by writing a one sentence characterization of the reader's probable difficulty, for instance, "the reader needs a definition of this concept." Students were not provided with any instruction on diagnosing readers' problems but were asked to phrase their characterizations in a way that would allow another person in the class to understand the problem they thought the reader would have.

Next, students were given the second half of the lesson: a think-aloud protocol transcript of a member of the audience reading and attempting to understand the text. Students were asked to read the protocol with the goal of using it to help them detect and diagnose additional problems that were made evident by the reader. During their second pass, students were encouraged to pay special attention to those problems they missed on their first pass, additional problems the reader feedback helped them to see and describe. To summarize, the teaching method for each of the ten lessons had two parts:

***Part One***

1. read the draft text;
2. predict the reader's problems with the text (detection phase);
3. characterize the reader's potential problems with the text (diagnosis phase).

***Part Two***

1. read the think-aloud protocol transcript;
2. use the reader's responses to identify additional problems (detection phase);
3. use the reader's responses to characterize the additional problems (diagnosis phase).

During their first pass, students predicted and diagnosed problems on the basis of their intuitions; during their second pass, they detected and diagnosed new problems revealed by the readers who were attempting to understand the texts. Students worked in this manner through each of the ten lessons. (At no time during the course of the ten lessons did students receive feedback about the quality of their performance, nor did the teacher discuss the teaching method.) Students were told they would receive feedback at the end of the series of lessons. The aim was to determine if analyzing readers' responses (via the think-aloud protocol examples) would influence writers' abilities to anticipate readers' problems—without the need of a teacher's interpretation and explicit instruction.

***Teaching Procedure for the Control Classes***

Writers in the control classes were taught to anticipate the reader's needs through a variety of audience-analysis heuristics and peer-response methods, including peer critiquing, role playing, and purpose-oriented audience pedagogies. Students were taught using mainly small-group exercises. Once or twice a week, they critiqued one another's papers, worked in collaborative groups, role played, or examined a variety of "good" and "bad" model texts of the sort written in the professions, for example, reports, proposals, memos, and so on. Teachers were rigorous in providing students with detailed feedback about how well their papers reflected the needs of the audience. All

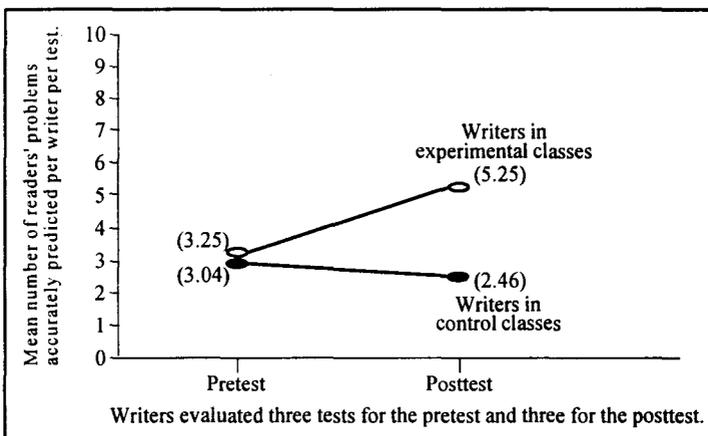
assignments in the course were written for audiences other than the teacher. For each paper, students were advised to elicit comments from members of their intended audience. Teachers reported that most students participated fully in classroom activities. Teachers who taught the course had been using this combination of methods over a number of years.

### *Pretest and Posttest Procedure*

Writers in both experimental and control classes were tested using the six texts described above. The six testing texts were divided into two groups: three texts labeled A and three labeled B. Half of the participants, both experimental and control, were pretested on the A texts and posttested on the B texts. The other half were pretested on the B texts and posttested on the A texts. For each of the six texts, writers were asked to predict the location and nature of problems that a freshman reader might have with the text. In this way, the testing procedure mirrored the *detection* and *diagnosis* phases of the reader-protocol method. The essential difference was that neither group (experimental nor control) was provided with reading protocols during pretest or posttest.

## RESULTS

I compared experimental and control classes to evaluate their



**Figure 2.** The effect of reader-protocol testing on writers' ability to accurately anticipate readers' problems.

improvement in *accurately* predicting reader problems. An accurate prediction is one in which writers predict problems that readers actually had as measured by the reading protocols collected from the freshmen. Results, graphed in **Figure 2**, indicate that writers in the experimental classes improved dramatically from the pretest to the posttest, increasing in accuracy by 62%, while writers in the control classes remained essentially unchanged. Analysis of variance indicates that the experimental classes' gain scores were significantly greater ( $F = 26.037$ ;  $df = 1, 8$ ;  $p \leq .001$ ) than those of the control classes. In addition, there was no significant difference between the accuracy of experimental and control classes' pretest scores ( $F = .685$ ;  $df = 1, 8$ ;  $p = .432$ ). At posttest, writers in the control classes tended to decline somewhat, but not significantly so.

In addition to asking if writers improved in their accuracy of predictions, I was concerned with whether writers in the experimental and control classes had changed in their ability to differentiate actual reader problems from non-problems. One can imagine that a teaching method of this sort could make writers hypersensitive to text problems, leading them to say that everything is a problem. Consequently, it was important to determine whether problems writers predicted were, in fact, problems for readers.

An analogous situation is that of judging the quality of a book reviewer. A reader evaluating a book reviewer's performance would want to know more than the reviewer recommends a high proportion of the good books on the *New York Times* book list that were published during the year. If that were enough, we could all become good book reviewers simply by recommending every book that is published. We would praise every good book, but we could hardly be described as discriminating. We would display no sensitivity to the differences between good and bad books. What we want in a good book reviewer, then, is someone who praises good books but not bad ones. Similarly, it is not enough that writers accurately predict a high proportion of readers' problems. If that were enough, we could all become good predictors of readers' problems simply by identifying every text unit as problematic. Writers must be able to discriminate between text units that are and are not potential problems for most readers.

The desirable situation is one in which the writer identifies a higher proportion of potential reader difficulties in text units that *are* problematic than in text units that *are not*. In other words, we want the probability that a writer says that an element is problematic when in fact it is (the probability of a hit) to be greater than the probability of saying that an element is problematic

when it is not (the probability of a false alarm).

To analyze the relationship between writers' accurate and inaccurate predictions, I used "signal detection analysis" because it takes into account both the hit rate and the false alarm rate in evaluating the sensitivity of performance. An ANOVA on the gain scores shows that writers in experimental classes were not just "problem happy" but, in fact, had increased in their ability to differentiate problems from non-problems significantly more ( $F = 8.752$ ;  $df = 1, 8$ ;  $p \leq .018$ ) than had control writers.

These results suggest that writers taught with the reader-protocol method improved significantly in their ability to anticipate readers' needs and that the sensitivities writers developed are helpful in discriminating problems that readers actually have. These findings raised the question, "What sorts of readers' problems are writers getting better at anticipating?"

### *Writers' Diagnoses of Readers' Problems*

To answer this question, I evaluated the kinds of problems writers diagnosed and how the diagnoses differed from pretest to posttest. A sample of 100 diagnoses from the pretest and 100 from the posttest for both the experimental and control groups (for a total of 400 diagnoses) was evaluated from the over 2,800 diagnoses writers made. My goal was to determine whether the reader-protocol method affected the kinds of problems writers noticed from pretest to posttest.

To capture the various dimensions of writers' diagnoses, I created three coding schemes. My goal in employing three schemes in the analysis was to view the data in complementary ways. All 400 diagnoses were coded three times, once for each scheme.

#### 1. Problems Focused on the Reader, Self, or Text

- a. *Reader-focused diagnosis*—characterized by an explicit mention of the reader, as in "A freshman student may not have an understanding of photography and would probably get confused when reading the analogy between photography and holography."
- b. *Self-focused diagnosis*—characterized by the use of "I," for example, "How do I know what causes 'inward growing grins?'"

From my point of view, 'inward growing grins' are ridiculous. This text doesn't have a clear purpose to me."

- c. *Text-focused diagnosis*—characterized by their similarity to stylistic maxims and grammatical rules, as in "You should always define concepts the first time they are mentioned in a paragraph."

## 2. Problems of Commission versus Omission

- a. *Problems of commission*—created by what the text says, that is, problems caused by what is on the page. This category includes anything that can be diagnosed by looking at the text as written and calling attention to a potential problem it creates, such as, "This is written in passive voice and that is bad."
- b. *Problems of omission*—caused not by what is on the page, but by what is not on the page. This category involves diagnoses of potential problems that are caused by what the text is missing. Problems of omission are brought about by missing content, underdeveloped ideas, gaps in logic, incomplete procedures, or non-existent transitions. For example, one student diagnosed a problem in this way: "First it talks about 'vitreous humor' and tries to define it, and then it switches to things like what causes blindness. I think something needs to be 'put in' to help the reader see why things like blindness and diabetes are related to vitreous humor somehow. Something is left out."

## 3. Problems at the Global or Local Level

- a. *Global diagnosis*—characterized by a description of a reader's problem conveying that the writer is attending to text elements larger than a sentence, such as, "This description of holography doesn't flow from one idea to the next. It seems kind of jumbled up like those 'jumbled laser beams.'"

- b. *Local diagnosis*—characterized by a description of a reader’s problem that shows the writer is focussing on text units of a sentence or smaller, as in “This word ‘nozzle’ sounds easy but it’s not. Is it like a hose? They should pick a better word.”

		Pretest	Posttest	Change
<b>Writers in experimental classes</b>				
N= 43 students				
	Reader	17%	22%	+16%
	Self	28%	43%	+5%
	Text	45%	24%	-21%
<b>Writers in control classes</b>				
N= 74 students				
	Reader	9%	13%	+4%
	Self	32%	35%	+4%
	Text	59%	52%	-7%

**Table 1.** Proportion of writers’ diagnoses of readers’ problems focused on reader, self, or text.

These coding schemes were derived from the literature on revision and the freshmen readers’ problems in comprehending the testing materials. Results from analyzing the diagnoses along these three dimensions—reader, self, or text; omission versus commission; and global or local—provided converging results. Based on the results of the coding for the 50 writers, I calculated the expected frequency of each diagnostic category for the 117 writers who participated. I used writers’ overall predictions to estimate the number of diagnoses writers made in each category.

**Table 1** displays the changes in percentage of reader-, self-, and text-focused diagnoses from pretest to posttest. At pretest, writers in both groups frequently made text-focused diagnoses. These results are not surprising, given the traditional emphases in English composition classes. However, at posttest, both groups reduced in their percentages of text-focused diagnoses, the control group by 7% and the experimental by 21%. Writers in both groups tended to move away from a focus on the text and paid more attention to themselves and to the reader.

It appears that both the reader-protocol method and the collaborative methods made students more aware of optional ways to diagnose text problems.

At posttest, writers in both experimental and control classes increased in their “self-focused” diagnoses. They more frequently monitored their own comprehension to predict what might trouble another reader, as the following diagnosis of “Vitreous Humor” suggests:

When I read the title “Vitreous Humor,” I thought it was talking about a type of new joke. Now that I read the whole thing, I still don’t understand it. And I bet others won’t figure out this part about people going blind from diabetes either.

In addition to using themselves as a model of the reader, writers in the experimental classes made more diagnoses in which they distinguished themselves from the reader with statements such as the following from a diagnosis of “Artificial Heart”:

I read about artificial hearts when William Shrader got his, and I think some guy from Louisville got one too. But a freshman who hasn’t read that story would never get this stuff. It’s too complicated and the “plutonium” reference is scary. What if the “metal carrier” comes open? Is that why they seem to croak-off so soon?

As **Table 1** shows, writers in the experimental classes showed the largest increases in reader-focused diagnoses. At posttest, writers taught with the reader-protocol method were much more prone than writers in control classes to make diagnoses such as the following based on “Artificial Heart”:

	Pretest	Posttest
<b>Writers in experimental classes</b>		
N= 43 students		
Omission	.84	3.76
Commission	3.84	3.61
Σ	4.68	7.37
<b>Writers in control classes</b>		
N= 74 students		
Omission	.65	.96
Commission	3.19	2.24
Σ	3.84	3.20

**Table 2.** Writers’ diagnoses of omission and commission: mean number per writer per text.

A freshman might take the phrase “a miniature nuclear furnace keeps an artificial heart beating” too literally. They might think doctors put radioactive stuff in your body. The writer is being too metaphorical for freshmen, too much like I’m a new hip reporter from *Discover* magazine.

An ANOVA on the increases in writers’ diagnoses from the reader’s perspective revealed a significantly greater increase for writers in experimental classes than for those in control classes ( $F = 26.133$ ;  $df = 1, 8$ ;  $p \leq .001$ ). These results suggest that the reader-protocol pedagogy heightened writers’ awareness of the audience more than did the conventional pedagogy.

**Table 2** summarizes writers’ diagnoses of omission and commission. Again, at pretest, writers attended closely to the text-as-written, spending most of their diagnostic activity describing errors of commission. However, at posttest, writers in the experimental classes increased in their diagnoses of how missing information might create problems for readers. An ANOVA indicated that the experimental classes’ shift toward diagnosing problems of omission is significantly greater than that of the control classes ( $F = 48.133$ ;  $df = 1, 8$ ;  $p \leq .001$ ).

At posttest, writers in the experimental classes seemed especially adept at perceiving gaps in the logic of the text or in detecting missing content. They made diagnoses such as the following:

By the time the reader gets to this idea, you have forgotten the main point. Something is missing here. There is a leap in what is said. The writer should restate the big picture of the flywheel idea. I do

	Pretest	Posttest
<b>Writers in experimental classes</b>		
N= 43 students		
Global	.98	3.76
Local	3.70	3.61
Σ	4.68	7.37
<b>Writers in control classes</b>		
N= 74 students		
Global	.92	.99
Local	2.92	2.21
Σ	3.84	3.20

**Table 3.** Number of local and global diagnoses: mean number per writer per text.

not know why I am being told this stuff—I need the purpose said again, if it was ever said to start with. [diagnosis of “Flywheels”]

What this eyeball passage needs is a diagram of an eye. Why would you write this without a picture? The writer must have forgot it or else he doesn’t care if the reader knows what’s going on. [diagnosis of “Vitreous Humor”]

It is important to point out that writers in the control classes also made such diagnoses at posttest; the essential difference was that they made many fewer.

**Table 3** presents the relationship between global and local diagnoses for writers in the experimental classes and control classes. Previous research in revision would predict the pretest results; that is, writers in both groups started out with a tendency to focus on local problems. Local diagnoses, as one would expect, concentrated on diction and style. By the posttest, the experimental group increased significantly more than the control group in the number of global diagnoses they made ( $F = 38.4$ ;  $df = 1, 8$ ;  $p \leq .001$ ). Most writers’ global diagnoses concerned issues of coherence, logic, and organization.

This finding suggests that the reader-protocol pedagogy helped writers in the experimental classes to perceive more problems at the global level of the text, an important advantage for initiating effective revision activity.

## Discussion

When writers in experimental classes are compared to those in control classes, it appears that the reader-protocol method enabled writers in experimental classes to:

- better diagnose problems from the reader’s point of view,
- become more sensitive to problems caused by omissions, and
- increase their awareness of problems at the global-level of the text.

Taken together, the various analyses show that the reader-protocol method improved writers’ ability to anticipate readers’ needs. The results of this study provide strong empirical evidence that the reader-protocol method helped increase writers’ perceptual knowledge by teaching them to see and hear the audience as readers.

This study suggests that document designers who engage in usability

testing may improve their skills in detecting and diagnosing readers' potential problems with functional texts. In particular, the results suggest that there may be important cognitive advantages for writers who employ reader-focused methods to guide text revision (for a discussion of other reader-focused methods, see Schriver, 1989). Writers in this study were not only able to increase their ability to notice certain kinds of problems, but they were able to transfer their knowledge of audience from one domain, which in this case was a form of instructional text, namely computer manuals, to another genre, which in this case was expository science text. Document designers can conclude then that to gain the most from usability testing, they should attempt to consolidate what they observe during testing, asking "What have we learned from this reader and this rhetorical situation that can be used to guide planning and revision of other texts?" When document designers are able to answer this question for themselves, they will recognize that usability testing is not an end in itself, but can be an opportunity for building a model of the ways readers engage with functional texts.

## References

- Schriver, K. A. (1984). *Revising computer documentation for comprehension: Ten lessons in protocol-aided revision*. (CDC Tech. Rep. No. 14). Pittsburgh, PA: Carnegie Mellon University, Communications Design Center.
- Schriver, K. A. (1989). Evaluating text quality: The continuum from text-focused to reader-focused methods. *IEEE Transactions in Professional Communication*, 32(4), 238-255.
- Schriver, K. A. (1991). Plain language through protocol-aided revision. In E. R. Steinberg (ed.), *Plain language: Principles and practice*, 148-172. Detroit, MI: Wayne State University Press.
- Schriver, K. A. (1992). Teaching writers to anticipate readers' needs: A classroom-evaluated pedagogy. *Written Communication*, 9(2), 179-208.
- Swaney, J. H., Janik, C. J., Bond, S. J., & Hayes, J. R. (1981). *Editing for comprehension: Improving the process through reading protocols*. (Document Design Project Tech. Rep. No. 14). Pittsburgh, PA: Carnegie Mellon University, Communications Design Center. (Reprinted in E. R. Steinberg, Ed., 1991, *Plain language: Principles and practice*, 173-203. Detroit, MI: Wayne State University Press.)

Thibadeau, R., Just, M., & Carpenter, P. (1982). A model of the time course and content of reading. *Cognitive Science*, 6, 157–203.

