
**Discourse Studies
in Composition**

Edited by

Ellen Barton

Wayne State University

Gail Stygall

University of Washington

 HAMPTON PRESS, INC.
GRESSKILL, NEW JERSEY

Swales, John M. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press, 1990.

Wertsch, James V. *The Concept of Activity in Soviet Psychology*. Armonk, NY: M. E. Sharpe, 1981.

_____. *Voices of the Mind: A Sociocultural Approach to Mediated Action*. Cambridge, MA: Harvard University Press, 1991.

CHAPTER 13

Evaluating Professional Discourse: How Well Does It Work For Real Readers?

Davida Chamey

University of Texas at Austin

Discourse analysis is essential to the study of technical and professional communication. By analyzing workplace and public discourse, scholars investigate what genres and rhetorical strategies characterize printed and electronic documents, how writers produce and disseminate them, and how readers use and respond to them. The study of such discourse is important because of the high stakes involved; the goal is not simply to understand how this discourse functions but to improve it, to make legal and technical documents more accessible to the public, to prevent disasters like the explosion of the shuttle Challenger, to promote ethical business practices, to improve workplace conditions for professional writers, and to help students acquire the skills and critical capacities to communicate effectively on the job and in public life.

An important branch of research in technical and professional writing, document design research, focuses on ways to present information to readers as effectively as possible. For a long time, "effectiveness" was equated with a strongly universalist notion of "clarity." If only the text were clear enough, the story went, any writer's ideas would transfer smoothly, accurately, and quickly to any reader. This

approach assumed that ideas are perfectly well-defined and self-contained in the writer's mind and that only problems of phrasing or presentation interfere with perfect transmission of the same ideas to the reader. This "window-pane" view of language was shared both by the technical writers and by many scientists and professionals. It had far-reaching implications for how people perceived the roles of writers and readers in making meaning. Scientific and technical writing was often perceived as mechanical and uncreative. Writers were encouraged to use impersonal formulations and to avoid any figurative expressions that might distract from the essential content of the message; scientific discourse was assumed to be expository, presenting facts about the world, rather than argumentative. Slobodkin, a professor of ecology and evolution, illustrates this view in his critique of the near-literary writing style of Stephen Jay Gould, the paleontologist and popular science writer: "in most scientific prose, the author strives for clarity in the dual sense of expository simplicity and in making oneself transparent so that the empirical world is visible through the text but the peculiarities of the author are invisible" (503). The role of readers in this view was also extremely limited; their major task was assumed to be comprehension and integration of information with prior knowledge. The differences among readers, their purposes for reading (beyond literal comprehension), and the extent to which readers actively construct the meaning of any text were largely ignored.

The quest for clarity led document design researchers to investigate whether some textual structures are inherently less clear, or harder to understand, than others. They also hoped to find alternatives that would reliably make texts easier to read, understand, use, and remember. With methods drawn from psycholinguistics and reading development, researchers produced an array of guidelines alerting writers to textual features that frequently cause problems for readers, including syntactic structures such as double negatives and implicatures, typographic features such as sans serif fonts and upper case, and organizational structures, such as conditional sequencing. Helpful features were also identified, such as headings, white space, hedges. (For guidelines, see Felker, Pickering, Charrow, Holland, and Redish; for recent studies, see Crismore and Vandepopple, Hartley and Sydes, Hartley and Trueman.)

Although the guidelines were often well-founded, their ultimate usefulness began to be challenged. Wright ("Writing Technical" and "Quality"), a leading document design researcher, argued that as guidelines proliferated, writers were less and less likely to remember to use them or be able to apply them to good effect. (For a similar critique and a thorough history of document design, see Schriver.) Guidelines were often stated in too general a form to help writers

solve a specific problem (e.g., "Know your reader"). One guideline might conflict with another, such as "Be concise" and "Be clear." Some guidelines were so specific that they turned into arbitrary and mechanical prescriptions (e.g., "Keep sentences shorter than 20 words"). Even worse, a writer who followed sensible guidelines to develop a document still had little evidence that readers could easily comprehend it.

The idea that text could be made universally readable by following certain prescriptions was taken to an unfortunate extreme with the development of readability formulas. Readability formulas, such as those available with many word processing programs, provide a numerical calculation of the "grade level" of education a reader is assumed to need to understand a text. The formulas are based on the number of words in a sentence and the length of words in syllables. Because such formulas are easy and inexpensive to apply, they proliferated even though their validity is highly dubious. In his extensive critique, Duffy points out that readability formulas were not originally intended to guide writing, but only to assess finished text. So given a set of storybooks, the formulas may help teachers decide which to use for sixth graders. The problem arose when the formulas were taken as guidelines for writing. Duffy found that revising a text to improve its readability score did not make it more comprehensible. Other factors, such as the complexity of the topic and the readers' purposes for reading, ultimately outweigh the effects of word choice and sentence length.

Discomfort with guidelines and formulas, interest in the rhetorical nature of technical and scientific discourse, and attention to the psychological processes of reading and writing brought the window-pane view of language under severe challenge. Guidelines are still considered helpful, especially when they are grounded in research and viewed as heuristics rather than formulas. But in the past 20 years, the question of what it means for a document to be "effective" has significantly broadened from studying textual features in isolation to studying actual situations in which texts are produced and used. Some scholars take an historical angle, analyzing specific documents, tracing the conditions under which they were produced, and assessing their role in decision-making processes, to explain why they might have succeeded or failed. For example, the explosion of the shuttle Challenger led to several insightful analyses of the communications among the engineers and managers responsible for the launch (see especially Herndl, Fennell, and Miller; Winsor, "Communication," "The Construction"). In contrast, usability testing, the approach I focus on here, looks forward toward meeting an upcoming communicative need, providing a way for writers with a document still under

development to identify what aspects may cause problems for readers and assess whether the text is working well enough to satisfy readers' needs. Although they are most often used for functional documents such as manuals, brochures, forms, and legal documents, the techniques of usability testing also can be applied to many forms of communication—and may be adapted for writing classrooms. In what follows, I outline the techniques of usability testing, focusing on a method adopted from psychologists called "thinking aloud." I illustrate these techniques with a case study of the usability of two versions of a form. Finally, I discuss broader applications of the techniques of usability testing and thinking aloud for exploring how readers respond to other kinds of discourse.

OVERVIEW OF USABILITY TESTING

Usability testing centers on people, representatives of the actual intended audience, who are observed reading and using documents as they would normally do to accomplish a goal. Usability testing allows writers to identify areas where readers have difficulty, compare alternative versions of a document, and assess the effectiveness of revisions.

Usability testing for documents grows out of the discipline known as human factors research (or ergonomics) that emerged during and after World War II. Human factors research draws on theories and methods from psychology, physiology, and engineering to encourage the design of products and procedures that are better suited to human capabilities. Studying the usability of documents emerged as an important strategy for businesses and public agencies in the late 1970s and early 1980s. A few highly publicized incidents in the 1980s underscored the value of document design and usability testing. In 1983, Coleco Industries, one of the first companies to market a home computer, suffered huge financial losses when hundreds of its Adam computers were returned as defective. In fact, only about 5% of the returned computers had defective hardware or software. The problem was that the user manuals were so confusing and incomplete that purchasers (even experienced computer users) believed the computers were defective. They could not perform basic tasks and did not know to avoid practices that could erase their disks (Davis). Although this incident illustrated the costs of poor documentation, another incident illustrated the importance of careful testing. In 1986, the Internal Revenue Service released a revised version of the W-4 form that workers use to tell employers how much tax

to withhold from their paychecks. The new form was a disaster. It was so complex and difficult to complete that taxpayers were in revolt. Even professional tax consultants failed to produce the right answers. One of the problems in the development of the W-4 form was sloppy usability testing; it was tested on IRS clerical workers instead of typical taxpayers (Gutfeld).

Since then, usability testing has become standard practice in many public agencies and high-technology industries, some of which have in-house usability labs to test equipment and software, as well as written materials, such as brochures, forms, online help, Web sites, and computer documentation (Nielsen, "Usability Laboratories"). Popular college-level textbooks for technical writing have begun to include instruction in usability testing (Anderson, Barker). For discussions of various forms of usability testing, see Wright ("Writing"), Schumacher and Waller, and Rubin. For more practical advice on how to conduct usability tests, see Nielsen ("Usability Engineering" and especially his Web site at <http://www.useit.com/>).

In a usability test, members of the intended audience use a draft of a document in a naturalistic way, such as carrying out instructions, using the draft to solve a hypothetical problem, finding relevant passages in a reference work, or learning and remembering some information. The document tested need not be a final draft, but it should look like one, so that the users will treat it seriously. In industry settings, usability testing often follows an internal process of feedback from subject matter experts and revision; but document designers have also found ways to test early plans for a document, for example, by testing alternative tables of contents. Participants should be recruited from the intended audience—or be as much like audience members as possible. They should be assured that the goal is to assess the quality of the document, not to test their skill or intelligence, even though their performance provides evidence of the text's success or failure. The participants' actions may be videotaped or simply observed and described in notes (or scores on a checklist). If the test involves a computer, the computer may often record some of the user's actions. Ideally, the setting allows the tester to observe from a distance, to avoid the temptation to assist the user or to show surprise. Participants may also answer questionnaires that assess their attitudes toward the document, their ability to solve problems, or their memory for key points. After the tests are completed, participants may also be interviewed in a more open-ended manner.

Watching what users do provides rich feedback on the effectiveness of a document and where it should be revised. But quantitative analysis of their performance can also be useful for comparing alternative versions or deciding how serious a problem may be. Several

aspects of the user's performance can be measured quantitatively: the time they need to complete a task or subtask, the number and kind of errors they make, the type and amount of information they recall, the strength of their preferences. For some purposes, more detailed analyses may be performed. van der Meij describes a method for categorizing each unit of information in a computer manual and recording whether users skip it, read it, or try to act on it in an appropriate way.

As in any empirical study, making use of qualitative and quantitative results requires negotiation and interpretation. There is usually no independent criterion for determining when a document is understandable enough or easy enough to use. At some point, writers have to decide what degree of confusion is tolerable. It is usually impractical to insist on revising and retesting the document until all participants score 100% on all tests. The criterion should be based on the purpose of the document and the goals of the writers. A high criterion should be set in high-risk situations, such as operating and maintaining a nuclear power plant, but a lower standard may suffice for assembling a model airplane. Using the results of a usability test to decide what to revise and how also requires considerable interpretation and sometimes outright guesswork.

OVERVIEW OF THE THINK-ALOUD METHOD

Conventional usability testing provides valuable information about how quickly and accurately users can read a text, find information or use it to achieve some goal. Usability tests are also a good way to survey users about their attitudes. But the conventional methods of usability testing have limitations. Users often make mistakes, take unexpected actions, reread, become annoyed, spend time figuring out what to do. These actions may be reflected in the time they spend or in the number of mistakes they make. But these results may not reveal why users have trouble or what might work better. Another technique, asking users to think aloud as they read and use a document, reveals more of what is happening and aids in the analysis of problems. For example, thinking aloud may reveal whether users made a mistake because they added some numbers incorrectly or because they were following the wrong step of the instructions. Think-aloud studies also identify additional trouble spots. For example, users might spend a great deal of time and effort re-reading and puzzling over a question before finally answering it correctly. If writers look only at whether the answers are correct, they might assume that no changes are needed.

The think-aloud method was developed by cognitive psychologists to make some thinking processes available for observation and analysis. This method has been used to study the processes people call on as they perform skills like reading, solving math problems, and playing chess. The think-aloud method has always generated a certain amount of controversy. Some critics believe that it promises direct observation of all mental activity, which it does not and cannot. Some are concerned that thinking aloud in itself changes the processes that people use (for a recent study investigating this concern see Janssen, van Waes, and van den Bergh). Other critics in composition are resistant to process research because of ideological preferences (see Charney, "From Logocentrism"). Yet, think-alouds have led to many important insights into reading and writing processes. For recent discussions of think-alouds in reading and writing research, see Kucan and Beck, Pressley and Afflerbach, and Smagorinsky. For background on the psychological theory underlying think-alouds, see Ericsson and Simon.

In a think-aloud usability study, participants are asked to say aloud whatever occurs to them while carrying out the task, reading aloud instructions or error messages, saying aloud what they are writing, and saying aloud whatever thoughts occur to them as they decide what to do. Many people already mumble comments to themselves when they read or carry out some task; in a think-aloud session, they are simply asked to raise the volume to an audible volume. Participants should not be asked to describe or explain what they are doing or to focus their comments on any particular aspect of the task—these instructions would be likely to change the participants' behavior, when the goal is simply to observe what they would normally do. (See sample instructions to participants in Fig. 13.1.) Participants are often given an opportunity to practice thinking aloud by reading a brief passage or solving a problem before beginning the usability test. Although participants vary greatly in the amount they say aloud and while thinking aloud may slow them down, few have any difficulty with the technique. The thinking aloud comments are tape-recorded and transcribed for analysis. The transcription of a think-aloud is referred to as a protocol.

In order to analyze the participants' comments, the transcripts, or protocols, are segmented into clauses (usually t-units) or episodes. Then these segments may be categorized according to a coding scheme and analyzed quantitatively and qualitatively. Figure 13.2 shows a coding scheme that I have used in studies of academic discourse, with graduate students and faculty members thinking aloud while reading research articles in their fields (Charney, "A Study"; Paul and Charney). The comprehension categories reflect the

We want to find out how easy these instructions are to understand by seeing what happens when people read them and try to carry them out. It will help us revise these instructions if we can see where you become confused, reread, or try to figure out what they are doing. In order to record something of your process, we will ask you to TALK ALOUD as you read and follow these instructions.

What we mean by "talk aloud" is that we want you to say out loud everything that is going through your head as you read the instructions and carry them out. Many people mumble to themselves as they do things anyway. We just want you to say it louder.

Remember that we want you to read and carry out the instructions as normally as possible. It may take you a little longer to do this task while talking aloud, but it should not change what you do. We're NOT asking you to describe what you're doing or to explain the material to us. Just read and react to it as you would in order to carry out the steps.

It is most important that you keep talking. If you are silent for any length of time, we will prompt you by saying "please talk." Do you understand what we want you to do?

To practice the think-aloud technique, we would like you to try thinking aloud as you divide two numbers without using pen and paper. Divide 5181 by 24.

Figure 13.1. Instructions to participants in a think-aloud study

participants' efforts to anticipate content, to clear up confusions (by re-reading or drawing inferences), and to monitor their understanding. Categories like these had been used in studies of children's reading processes (Bereiter and Bird; Olson, Duffy, and Mack). Relatively few think-aloud studies had looked at adults reading for different purposes, or with different types of texts, such as argument, so previous studies had nothing corresponding to my categories for evaluation and text structure. I created these categories after reading the protocols themselves and attempting to sort the comments—they reflected readers' efforts to come to an independent stance on the issue. This scheme was very productive for revealing how readers engage the text; I found that faculty members and graduate students used similar strategies to comprehend the text and made about the same average number of comprehension comments, but that faculty members went on to make about twice as many evaluative comments as graduate students. The categories in this scheme are not definitive. The categories in a given study to some extent depend on the

COMPREHENSION PROCESSES

- Rereading a verbatim repetition of text already read
- Paraphrase a restatement or summary of the ideas in the text ("so when we fail to explain something by natural selection then we just ignore it")
- Inference a conclusion or implication warranted by the text but not stated there explicitly ("in other words, we are missing the point"), or a problem-solving episode to figure out what the text means ("neotenic" means it's retained in an early stage of growth and development . . .)
- Prediction a reference to or neutral comment about earlier information in the text, a prediction about what the text will be about, or a confirmation or withdrawal of an earlier prediction ("that's what I thought he was going to conclude")
- Metacomment a comment about the reader's own understanding or lack of understanding ("OK, got it," "unfamiliar term"); expletives or ambiguous comments ("OK," "Oh Lord"); a comment about a plan for how to recover from loss of comprehension ("better look at that again")

EVALUATIONS OF CONTENT, ARGUMENTS, AND CLAIMS

- Validity Judgment an explicit assessment of the truth or probability of an assertion ("OK, I can live with that," "Yes, I believe that," "that's not true," "maybe," "bullshit")
- Value Judgment an explicit evaluation of an assertion's interest ("well that's not surprising"), importance ("so what?") "that's a good point"), or relevance ("Oh this is a very orthodox view nowadays," "a rather poor example, at least not a familiar one")
- Reference to Prior Knowledge new information introduced by the reader including opinions, interpretations, examples, criteria, or associations about the text, authors, subject matter, etc. ("this is a very deterministic argument," "I can probably think of some examples here," "I wish I knew a little more of Aztec cosmology," "that's overly simplistic," "yeah, but how do you test that?")

REFERENCES TO TEXT STRUCTURE OR GENRE

- Intratextual A comment explicitly identifying or referring to a structure within connected prose, such as the abstract, introduction, conclusion, summary, title, examples, wording, definitions ("this is really sort of a final discussion," "And

Figure 13.2. Coding scheme for thinking aloud comments during reading

here's the quote from David's article," "and this is still going on parenthetically?")

- Extratextual
structure *outside* the connected prose, such as works cited, tables or figures, the journal citation line ("what's the date on this?" "any more figures?")
- Holistic
A characterization or comment on the text as a whole, its genre, goal, aim or approach ("this is an attack," "OK it's a pedagogic piece")

GLOBAL METACOMMENTS

- Metacomments
Comments on reader's normal or habitual behavior ("I'd normally go look that up"), reader's immediate behavior or circumstances ("this is a bad copy," "I should xerox that actually," "I'm drifting from it"), or current reading strategy ("OK . . . I'm just going to run through this paper")

Figure 13.2. Coding scheme for thinking aloud comments during reading (cont.d)

reading situation (the text, the field, the purpose for reading) and on the research question of the study. For example, Wyatt and colleagues created a much more detailed inventory of reading behaviors that social scientists use when reading scholarly articles.

Researchers have used various approaches to segment and code protocols. In their early work, Flower, Hayes, and Swarts sometimes coded on the basis of clauses and sometimes used looser "episodes," defined as "units of concentration." Bracewell and Breleux argue for a more fine-grained, semantic approach to coding that works at the level of propositions. The rigor of the coding to some extent depends on the purpose of the study. A fully detailed analysis may not be necessary if the goal is simply to assess the difficulty of a text and identify places to revise. When the goal is to categorize every comment, in order to look at how readers spend their time or what kinds of comments they make most frequently, then it is important to provide some kind of check on the coherence and reliability of the coding scheme. An independent rater may be asked to use the coding scheme to categorize all or some of the comments. Then the reliability, or consistency of agreement, between the raters can be calculated. As in any other kind of discourse analysis, the establishment of a

coding scheme is an effort at objectivity, an effort to establish a system of research procedures that can be checked for consistency of application. These efforts are then described and defended in the method sections of a research article, and thereby opened to scrutiny, challenge, reuse and refinement by the disciplinary community. For a wider discussion of scientific objectivity and its ideological implications, see my "Empiricism." For a discussion of objectivity in the coding of discourse, see Hillocks.

Think-aloud studies involve more effort than conventional usability tests, because of the need to transcribe and analyze the protocols. For this reason, thinking aloud is not used as often as other methods. But the value of the insight that think-alouds provide is great. Writers can more easily identify with readers when they see part of what readers are thinking. In fact, Schriver found that practice at diagnosing readers' problems from think-aloud protocols leads to lasting improvement in students' abilities to detect problems and take them more seriously.

CASE STUDY: DESIGNING AND TESTING A WORK ORDER FORM

To illustrate the value of user testing and think-aloud protocols for seeing the effects of textual choices on readers, I describe a project in which I worked with the staff and clients of a university printing and graphics service to redesign a work order form and to test whether the revised form was easier for clients to fill out than the original (for a more detailed discussion, see Charney, "Redesigning"). Clients of the university printing office used the form shown in Figure 13.3 to request a variety of services from typesetting and photography to photocopying, off-set printing, and binding. The office manager (who had designed the original form) wanted it revised because it was out of date (listing services that were no longer available and omitting some new ones) and because it was difficult for clients to complete accurately on their own.

In order to analyze the form and the discourse of the people who used it, I interviewed staff members to find out what the terms on the form meant (and how they were understood in different parts of the office), what kinds of information might appear in each space, what kinds of questions clients had about the form and what mistakes they made in filling it out, and what parts of the form no longer reflected the services available from the office. In addition to working to understand the form itself, I also needed to understand how it

functioned in the office, how it moved a printing job through the parts of the office, who handled it and what they did with it. Finally I dug through files of completed forms to get a sense for the range of requests and what kinds were most typical—if space were at a premium, it might make sense to use it to facilitate the most typical jobs.

I was surprised to learn through these interviews that most staff members were fairly satisfied with the form. They were used to it; they thought a few minor corrections would suffice. To me, it seemed obvious that the original form had serious problems in its layout and its language. The layout is a grid that imposes a false sense of uniformity on the options and obscures the different relations among them. For example, in the right-hand portion of the Camera section, clients are asked to indicate "Size" by checking a box under "100%" or "Enlarge/Reduce." If they choose the latter, then they are supposed to go on to fill in the next space, "Final Size % of Original," but this item is not grouped under "Size" with the others. Similarly, the options under "Stitching" all appear to be equivalent, even though "Upper Left," "Left Margin," and "Saddle Stitch" all describe methods of stapling, while "GBC," "Tape," "Velo," and "Perfect" describe different binding materials. A grid format also unnaturally restricts the size of the spaces: the space available is unrelated to the amount of information required. For example, the space allowed for specifying the color of ink is too small, when the client must write something like "Pantone 347." Specifying the paper stock requires considerable space for brand and weight information, so separate lines are provided for stock information below the grid. But this placement forces clients to match each stock line by letter code to the items in the grid above. The least functional parts of the form were the gray areas where staff members were supposed to keep track of which operations had been completed, but where space was too limited to do so. The space problems were well known to the staff; they handled it by attaching a separate page with ruled lines called the Additional Information Sheet (AIS).

The language of the form included constructions that psychologists had long noted as difficult for readers, such as the negative conditionals "Indicate Only if Not," and "If Other than 8-1/2 x 11." The purpose of the conditional is to establish the default, so that clients can leave these spaces blank if they want the standard size of paper and standard color of ink (black). But conditionals are notoriously hard to understand. Some terms were also ambiguous; for example, do "No. Pages One Side" and "No. of Pages Two Sides" refer to originals or copies? The form also included acronyms and technical terms that a general audience might not know such as "bleeds," "half-tones," "GBC," and "saddle-stitch." All these seemed candidates for revision.

REQUISITION Name: _____ Department: _____ Ext: _____

Job No. _____ Date Submitted _____

Date Required _____

Approved _____

PUBLICATIONS/CAMPUS PRINTING

Services Requested: Design Composition Key Art

Indicate Only if Not: 8 1/2 x 11 8 1/2 x 11 11 x 17

Quantity of Originals: _____

Quantity of Copies: _____

Color: Black CMYK Pantone

Ink: Black CMYK Pantone

Size: _____

Final Size % of Original: _____

Stitching: Upper Left Left Margin Saddle Stitch Velo Tape Perfect

Stock: (If other than 8 1/2 x 11, 20# Bond, White used)

Item	No. of Originals (Print)	No. of Copies (Print)	Color	Ink	Size	Final Size % of Original	Stitching	Stock	Quantity	Operator Initials	Date Printed	No. Used
A.												
B.												
C.												
D.												

Camera: Camera Xerox Reduction Requested AIS Required

Design: (1-081-20-7816) Composition (1-081-20-7814)

Key Art: (1-081-20-7817)

Copies: (1-081-20-7811)

Mailing: (1-081-20-7812)

Miscellaneous: (1-081-20-7813)

10 Cards: (1-081-20-7815)

Miscellaneous: (1-081-20-7880)

1st Class Postage (1-00000-1639)

3rd Class Postage (1-00000-1639)

TOTAL CHARGES \$ _____

Authorized By: _____ Budget No. _____

STATEMENT OF CHARGES

Design (1-081-20-7816) \$ _____

Composition (1-081-20-7814) \$ _____

Key Art (1-081-20-7817) \$ _____

Copies (1-081-20-7811) \$ _____

Mailing (1-081-20-7812) \$ _____

Miscellaneous (1-081-20-7813) \$ _____

10 Cards (1-081-20-7815) \$ _____

Miscellaneous (1-081-20-7880) \$ _____

1st Class Postage (1-00000-1639) \$ _____

3rd Class Postage (1-00000-1639) \$ _____

TOTAL CHARGES \$ _____

Authorized By: _____ Budget No. _____

DELIVERY DATA: Addressing/Mailing (Fill out mailing specification form)


1/78 PUB002

Figure 13.3

The staff art director and I worked together to create the new version of the form shown in Figure 13.4. We showed sketches of the new version to the staff and incorporated their suggestions. The biggest changes were replacing the grid with checkboxes and blanks and removing the gray tracking areas altogether. Mutually exclusive choices were more clearly signaled by adjoining checkboxes with "or." Horizontal ruled lines were used to separate functions or items. Wherever possible, we substituted descriptive terms for technical terms. For example, "GBC" and "Velo" were replaced with "Plastic Comb" and "Plastic Strip," and "No. Pages One Side" and "No. of Pages Two Sides" were replaced with "Single-sided" or "Duplex."

Intuitively, the new form seemed more attractive and easier to use than the original. However, because the staff seemed unconvinced that the original form posed serious problems for clients (especially their regulars), I decided to test whether the changes made a significant difference, either to regular clients or to new prospective clients. Ideally, regular clients would not have to struggle to learn to use the new form and prospective clients should easily understand their options. To show a change in usability, I needed to observe clients using both forms, to measure differences in how quickly and accurately they could fill the forms for the same printing job. I also planned to collect think-aloud protocols.

Deciding on the tasks the clients would perform required juggling several goals. I wanted to test all the parts of the form, but I also wanted the task to be naturalistic. I considered watching actual clients submitting real jobs, but most jobs did not use all of the parts of the form. Furthermore, observing clients using the form for real jobs would not allow direct comparison of the original and revised forms, because the jobs would vary so much in difficulty. For these reasons, I decided to construct four realistic but hypothetical printing jobs, each one fairly simple in itself, that together would require use of different parts of the form. One task called for ordering copies of a report that was to be bound in heavy covers with a plastic spiral binding. A second task called for camera work, making an enlarged print of a department logo using a photo-mechanical transfer. The third task called for printing and folding invitations, and printing reply cards and envelopes. The fourth task involved photocopying an article, making reductions so that each side of the final copy contained two pages of the original. In each case, participants were given a set of "originals" to be submitted to the printing office, a blank form, and some labeled samples of different paper stocks and inks. To convey details of the task without giving away crucial details (like technical terms), participants were also given facsimiles of the final products so they could compare their original to a model end-product. For exam-



WORK ORDER

Work Order No. _____ Date Submitted _____ Date Requested _____
 Client _____ Ext. _____ Department _____ Approved By _____
 Title of Job _____ Send Charges to Name _____ Department _____ Date Completed _____

See AIS Writing/Editing Design Photography Composition Art

Publication

See AIS Xerox Reduction (Reduce original to _____)
 Item A _____ No. of Originals _____ Quantity Required _____
 Finished File Size _____ Ink _____ Bleed Recut
 Stock _____ Single Sided or Duplex Stock Supplied

Item B _____ No. of Originals _____ Quantity Required _____
 Finished File Size _____ Ink _____ Bleed Recut
 Stock _____ Single Sided or Duplex Stock Supplied

Item C _____ No. of Originals _____ Quantity Required _____
 Finished File Size _____ Ink _____ Bleed Recut
 Stock _____ Single Sided or Duplex Stock Supplied

Camera

See AIS See Original Same Size Enlarge or Reduce Original to _____
 Item 1 _____ No. of Originals _____ 10 x 12 or 12 x 18 Print Size Line or Halftone PMT or PMT Reverse or Film Negative

Item 2 _____ No. of Originals _____ 10 x 12 or 12 x 18 Print Size Line or Halftone PMT or PMT Reverse or Film Negative

Item 3 _____ No. of Originals _____ 10 x 12 or 12 x 18 Print Size Line or Halftone PMT or PMT Reverse or Film Negative

Bindery

See AIS See Sample See Original See Sample
 Collating Collate Items (See AIS) Slip Sheet Insert (See AIS)
 Stapling Side Stitch 2 or 3 Saddle Stitch 2 or 3
 Upper Left (One Staple) Plastic Strip Perfect

Binding 1/2 1/2 x 1/2 1/3Z LTR
 Folding 1/2 BL/1/2 1/2 BL/1/2 LR Other (See AIS)

Finishing Trim Perforate Number (From _____ To _____)
 Die-Cut Drill Holes (No. of Holes _____) NCR
 Send wraps (_____ per page) Other (_____ per page)

Delivery

See AIS See AIS Deliver to Bldg. _____ Call Est. _____ for Pickup
 See Mail Specifications Hold for Mailing (Date to Bldg. _____)

Authorized Signature _____ Budget Number _____

Charges

Item # Description Unit Price Total
 1001 Photography 1.00 20.00
 1002 Writing/Editing 1.00 20.00
 1003 Design 1.00 20.00
 1004 Composition 1.00 20.00
 1005 Art 1.00 20.00
 1006 Xerox Reduction 1.00 20.00
 1007 Camera 1.00 20.00
 1008 Bindery 1.00 20.00
 1009 Delivery 1.00 20.00
 1010 TOTAL CHARGES

Figure 13.4

ple, for the first task, they were given a copy of a bound report and instructed to make theirs come out like it. The facsimiles actually increased the realism of the tasks, because clients often brought in examples to show how they wanted their product to look.

As an example, here are the instructions for the first task:

Your department occasionally publishes technical reports. You have been put in charge of getting 250 copies of the supplied report printed and bound with covers. The report entitled "Studies of Metaphor and Complex Analogies" is the report you wish to have copied. You are to fill out the form in such a way that your report will come out looking just like the technical report entitled "How Headings in Documents Can Mislead Readers." Make sure that the title page, the page of tables and the page of figures are left blank on the back. The reports must be ready for you to pick up next Monday.

To find representative users to participate in the study, I identified six members of the university staff who were regular clients and six others in similar positions who had not used the printing service yet but were likely to in the future. The regulars had used the form for a variety of purposes but none was familiar with all the services available. The number of participants, 12 in all, was small for a study in which I hoped to compare times and error rates quantitatively, but fairly large for a study with think-aloud protocols, which produce large amounts of data. The number of participants in most workplace studies is kept small because the standard of evidence is not as high as it would be in a more scholarly study. Even a little evidence of difficulty from users may be enough to justify revisions, even though there may not be enough evidence to say that users in general would consistently have that amount of trouble. All participants filled out forms for all four tasks; half the participants used the original form for all four tasks and half used the revised form. By measuring the time it took users to fill out the forms and identifying errors for four tasks, it would be possible to compare average performance across all four tasks.

The participants were asked to think-aloud while completing the forms, with instructions similar to those in Figure 13.1. Each participant came to a quiet room to complete the tasks; they were timed for each task and their think-aloud comments were taped. Each participant worked on the four tasks in a different order (determined by randomization), so that if they became quicker and faster with practice, this gain would not always be registered for the same task. After the participants completed all four tasks, I

asked for comments on both versions of the form. Not surprisingly, when I calculated the time needed for filling out the forms, the regulars were quicker than prospective clients; the regulars needed an average of 10.7 minutes per form, while the prospective clients took 14.7 minutes. More gratifying was finding that both regular and prospective clients filled out the new version more quickly than the original. Regulars who used the new form took an average of 4.5 minutes less per form than regulars using the original form—even though they were already more familiar with the original. Prospective clients needed about 3 minutes less if they used the new version than if they used the original.

More complex were the results of analyzing the number of mistakes. Mistakes included omitting information, checking the wrong item, or writing in erroneous information. On the positive side, regulars clearly made fewer mistakes on the new version (6.4 mistakes per form) than on the original version (9.8 mistakes). Considering that the regulars were familiar with the original form, this was a significant sign that the new version was an improvement. But having found the original form fairly intimidating on first sight, I expected prospective clients to have more difficulty filling it out than the new version. In fact, I found that prospective clients made more mistakes on the new form (10.7 mistakes) than on the original (8.8 mistakes).

This finding led me to look more closely at the kinds of mistakes clients made, because the printing office staff considered some kinds of mistakes more serious than others. The most common mistake is simply leaving off some necessary piece of information. But this kind of "omission error" is not that serious because it is easily noticed. For example, a print operator cannot proceed without knowing how many copies to make, so if this information is missing, the operator will contact the client. Similarly, if a client indicates that a set of reports should be stapled, the staff can easily infer that the pages must be collated, even if the collating space is blank. More serious are "misplaced entries," checking the wrong option or supplying information in the wrong place or "wrong entries," using the right place but entering the wrong information. In both cases, the printing staff may justifiably assume that the option was chosen deliberately and attempt to carry out the instructions as written. Misplaced entries are easier to detect because marking an irrelevant space often results in conflicting instructions. In contrast, a wrong entry may seem perfectly reasonable, even though it does not reflect what the client knows or wants. For example, if a client who wants a photograph reduced by one third writes in 30% instead of 70% under the "Final Size Percent of Original," the staff has no way to detect the mistake.

When I took the seriousness of the errors into account, I found that prospective clients made less serious errors on the new form, even if the absolute number of errors was not smaller. Using the new form, prospective clients who were confused were more likely to leave a space blank; using the original form, they were more likely to write something incorrect.

The comments from the think-aloud protocols supported the evidence that we had found effective solutions to some real problems in the original form. For example, many participants expressed confusion about the technical terms on the forms. The following comments from three participants indicate that they also found the negative conditionals in the original form very confusing, even if they ultimately figured them out. (Italics indicate a first reading of verbatim language from the form. My descriptive comments are in brackets.)

Indicate only if not . . . Only . . . I don't understand that . . . Indicate only if not . . . only if not 8-1/2 by 11 original? Oh, I see.

Indicate only if not . . . 8-1/2 by 11, 8-1/2 by 11, Black. I have no idea what that means . . . Indicate only if it's not these sizes. Oh I see.

Indicate only if not . . . black. It's not black [looking at the set of original pages]. Oh, the ink, the letters? Yeah. I'll leave that blank. If it's not 8-1/2 . . . , indicate it. OK

These comments illustrate several behaviors that indicate difficulty in comprehension, including multiple rereadings, explicit statements of confusion, questioning the text, inferring the meaning, and paraphrasing to produce a simpler syntax. The fact that several participants, including regulars, had trouble comprehending the negative conditionals suggested that it was worth removing them, even though most participants filled in these spaces correctly. The additional time it took participants to figure out these phrases surely contributed to the longer task completion times for the original form—but without the think-aloud protocols, it would have been impossible to say where the extra time was spent. On the new form, the negative conditionals were eliminated, but clients had to fill in the size of the paper and the color of ink on every job. Two of the experienced subjects were surprised at having to indicate this information. The cost of the extra writing does not seem very great, considering that regular clients completed the new form more quickly and with fewer mistakes than the original form.

Similarly, the protocols revealed problems caused by the grid that would not have been detected through time and error analyses.

In the following excerpt, the client (a regular) was trying to request a plastic spiral binding, called a GBC. Even after she finds the right term, the proximity of other options leads her to believe that she has to say more about where the page is to be bound:

Bindery—ooh. . . . I don't know what this squiggly thing is called. [Indicates plastic spiral binding. Asks for definitions of Velo, Perfect, GBC. Told GBC is a plastic spiral binding.] How is a normal person supposed to know that? God. OK. Ah, that was lucky. *Upper left.* Upper left what? The whole thing—I want the whole thing done. *Left margin.* Yes. 1, 2 and 3. I don't know what they're talking about. Left margin. Well let me first mark stitching. [Puts X under "Stitching"—which is incorrect as it refers to stapling.] OK. GBC—yes, I want that done. What do I put—"on all of them"? I'll just put an "x"—GBC. [Puts X under "GBC."] I don't know what Upper Left is. Oh I know, that's probably a staple in the upper left, probably. Well, whatever, I'll just put GBC. That was easy enough.

This comment illustrates some of the frustration that readers feel when they have trouble, although in this case, the participant has an unusually positive outlook. When Karen Schriver observed users trying out manuals for electronics equipment, she found that they often blamed themselves when they had difficulty, rather than assuming that the manuals were badly written. In contrast, this participant considers herself "lucky" and ends up thinking that this section was "easy enough." Both reactions are important—they help explain why attitude questionnaires may fail to detect problems. Participants may indicate satisfaction with a document only because they don't blame it for their troubles or don't realize that they are having more trouble than they should.

Two spaces in the Duplicating section of the original form caused clients more trouble than anything else on the form: "No. of Pages One Side" and "No. of Pages Two Sides." These spaces are intended to refer to the finished copy, to indicate the number of pages in one finished set that will have printing on one side or two sides. Filling in these spaces correctly requires a fairly abstract calculation from a mental image of the final product. In effect, the client must project how many sheets of paper will be used in each copy, a number that might be very different from the number of originals that the client is submitting. For the "Report Task" described above, the correct answer for Number of Pages One Side is three, because the title page, the page of tables and the page of figures are to be left blank on the back, and the Number of Pages Two Sides is four, because the

remaining seven pages of originals will be printed on four sheets. From the point of view of the printing office, this is important information so that the stockroom clerk can pull the right number of sheets and the cashier can charge the right amount. The client, on the other hand, just wants a way to tell the staff which sheets to leave blank on the back.

Every participant who used the original form made mistakes in these spaces for at least two of the three tasks that used this part of the form. So even without the think-aloud protocols, it was clear that these spaces needed to be changed. However, the think-aloud protocols indicated that clients (including regulars) gave quite wide-ranging interpretations to these labels.

Oh here's where No. Pages One Side becomes pertinent. Hum. I don't know how to fill this out. . . . all these pages [indicates originals] are one-sided but I want some of them to come out to be two-sided. Number of pages one side . . . that I want or that I have?

Now I've got to figure out a way to . . . OK, title page . . . [writes "Title page" under Items]. Title page—well, number of originals—one [writes 1 under No. of Originals]. I want 250 of them [writes 250 under Quantity Desired]. *Number of Pages One Side*—I guess that's 250 also. Yeah, because that's how many it would turn out to be.

I want ten originals [writes 10 under Originals]. I want 250 of each ten [writes 250 under Quantity Desired]. Yes. *Number of Pages One Side*. One, two—oh I think they all are. Number of pages on one side. Looks like every one of them. OK. Ten.

The wide variety of responses suggested that there was no easy way to convey what kind of calculation was needed. In the new version, we simply gave a choice between single-sided and duplex, leaving it to the staff to perform the calculation rather than the client.

In the end, I was pleased to find that the new version of the form was easier to use than the original version. But it was difficult to find that clients still made numerous errors on the new form. The testing led us to make additional changes to the new version, particularly in some of the terminology. It was also disappointing that the regular clients often told me that they preferred the original form—at a time when I was still collecting the evidence that showed that the new version was easier to use. The clients' attitude and the printing office staff's initial resistance to change gave me valuable insights into workplace dynamics.

CONCLUSION

By providing direct observation of readers using texts in naturalistic ways, usability testing is a valuable way to find out whether a document works for its intended audience and, if not, how to improve it. When the stakes are high, writers can continue to refine documents through a recursive cycle of writing, testing, revising, and testing, until they are confident that most readers will be able to use the document effectively in most situations. But apart from improving individual documents, usability testing also provides larger scale benefits for writers and for researchers interested in document analysis.

Usability testing provides an important complement to guidelines based on studies of the textual features. Usability testing very often confirms the wisdom of a guideline based on empirical research. For example, the finding that conditionals and negative expressions in the work order form caused problems for readers would not surprise psychologists who have worked with such constructions. But the quality of the data in usability testing is compelling enough to have special merits. Writers and their corporate managers are much more easily convinced that revision is necessary by a demonstration of real users having problems with a document than by the citation of guidelines based on experimental studies. Furthermore, the practice of conducting usability testing seems to confer benefits on writers, who may begin to anticipate readers' problems more accurately (Schrivver). Writers who plan usability tests have to attend to a rich array of variables that will make the testing conditions naturalistic; they must consider not only the features of the text but also the contexts in which documents are used, the people who use them and their varying purposes. Usability testing, then, can raise writers' awareness of the full rhetorical situation surrounding a document, the variables that guidelines, by their nature, leave unspecified. In the light of usability testing, the goal of discovering a set of textual features that will reliably confer universal "clarity" can be seen as improvable.

The techniques of usability testing and thinking aloud are not only valuable to writers on the job; they can also be used to investigate how rhetorical and linguistic strategies influence the reception of scientific articles and other professional genres. I used usability techniques to examine how grammar, style and elaboration choices in students' résumés influenced a job recruiter's willingness to grant an interview (Charney, Rayman, and Ferreira-Buckley). Similarly, Paul examined how readers in the scientific community respond to the components of article introductions identified by Swales and Najjar. The first part of her study was analytic; to see how these components

change as a new discipline emerges and matures, she analyzed the introductions in a dozen research articles written over a 15-year span by two important, active chaos scientists. Then using think-aloud methods, she analyzed which moves attracted the most comments from readers and how well the introductions succeeded in attracting and retaining readers' interest. By looking at the responses of real readers, scholars can learn more about the complex interactions among texts, readers, and writers.

REFERENCES

- Anderson, Paul. *Technical Writing: A Reader-Centered Approach*. 3rd ed. Fort Worth, TX: Harcourt Brace, 1995.
- Barker, Thomas. *Writing Software Documentation: A Task-Oriented Approach*. Boston: Allyn and Bacon, 1998.
- Bereiter, Carl, and Marlene Bird. "Use of Thinking Aloud in Identification and Teaching of Reading Comprehension Strategies." *Cognition and Instruction* 2 (1985): 131-56.
- Bracewell, Robert, and Alain Breleux. Substance and Romance in Analyzing Think-Aloud Protocols. In *Speaking About Writing: Reflections on Research Methodology*. Ed. Peter Smagorinsky. Thousand Oaks, CA: Sage, 1994. 55-88.
- Charney, Davida. "Redesigning and Testing a Work Order Form." *Information Design Journal* 4 (1984): 131-46.
- _____. A Study in Rhetorical Reading: How Evolutionists Read "The Spandrels of San Marco." In *Understanding Scientific Prose*. Ed. Jack Selzer. Madison: University of Wisconsin Press, 1993. 203-31.
- _____. "Empiricism is Not a Four-Letter Word." *College Composition and Communication* 47.4 (1996): 567-93.
- _____. "From Logocentrism to Ethocentrism: Historicizing Critiques of Writing Research." *Technical Communication Quarterly* 7.1 (1998): 9-32.
- _____, Jack Rayman, and Linda Ferreira-Buckley. "How Writing Quality Influences Readers' Judgments of Résumés in Business and Engineering." *Journal of Business and Technical Communication* 6 (1992): 38-74.
- Crismore, Avon, and Willem Vande Kopple. "Readers' Learning From Prose: The Effects of Hedges." *Written Communication* 5 (1988): 184-202.
- Davis, Bob. "Hundreds of Coleco's Adams are Returned as Defective; Firm Blames User Manuals." *Wall Street Journal* (30 November 1983): 14.
- Duffy, Thomas. Readability Formulas: What's the Use? In *Designing Usable Texts*. Eds. Thomas Duffy and Robert Waller. Orlando, FL: Academic Press, 1985. 113-43.
- Ericsson, K. Anders, and Herbert Simon. *Protocol Analysis: Verbal Reports as Data*. Bradford, MA: MIT Press, 1993.
- Felker, Daniel, Frances Pickering, Veda Charrow, V. Melissa Holland, and Janice Redish. *Guidelines for Document Designers*. Washington, DC: American Institutes for Research, 1981.
- Flower, Linda, John R. Hayes, and Heidi Swarts. Revising Functional Documents: The Scenario Principle. In *New Essays in Technical and Scientific Communication*. Eds. Paul V. Anderson, R. John Brockmann, and Carolyn Miller. Farmingdale, NY: Baywood, 1983. 71-89.
- Gutfeld, Rose. "Sad Returns of W-4: One Bureaucrat's 'Spirit' Becomes Another's 'Parody.'" *Wall Street Journal* (17 February 1987): 26.
- Hartley, James, and Matthew Sydes. "Which Layout Do You Prefer? An Analysis of Readers' Preferences for Different Typographic Layouts of Structured Abstracts." *Journal of Information Science* 22 (1996): 27-37.
- Hartley, James, and Mark Trueman. "The Effects of Changes in Layout and Changes in Wording on Preferences for Instructional Text." *Visible Language* 15 (1981): 13-31.
- Hernl, Carl, Barbara Fennell, and Carolyn Miller. Understanding Failures in Organizational Discourse: The Accident at Three Mile Island and the Shuttle Challenger Disaster. In *Textual Dynamics of the Professions: Historical and Contemporary Studies of Writing in Professional Communities*. Eds. Charles Bazerman and James Paradis. Madison: University of Wisconsin Press, 1991. 279-305.
- Hillocks, George, Jr. Interpreting and Counting: Objectivity in Discourse Analysis. In *Speaking About Writing: Reflections on Research Methodology*. Ed. Peter Smagorinsky. Thousand Oaks, CA: Sage, 1994. 185-204.
- Janssen, Daniel, Luuk van Waes, and Huub van den Bergh. Effects of Thinking Aloud on Writing Processes. In *The Science of Writing: Theories, Methods, Individual Differences, and Applications*. Eds. C. Michael Levy and Sarah Ransdell. Mahwah, NJ: Lawrence Erlbaum, 1996. 233-50.
- Kucan, Linda, and Isabel Beck. "Thinking Aloud and Reading Comprehension Research: Inquiry, Instruction, and Social Interaction." *Review of Educational Research* 67 (1997): 271-99.
- Nielsen, Jakob. *Usability Engineering*. Boston, MA: Academic Press, 1993.

- _____. "Usability Laboratories: A 1994 Survey." 11 Jan. 1999 (<http://www.useit.com/papers/uselabs.html>).
- Olson, Gary, Susan Duffy, and Robert Mack. *Thinking-Out-Loud as a Method for Studying Real-Time Comprehension Processes. In New Methods in Reading Comprehension Research*. Eds. David Kieras and Marcel Just. Hillsdale, NJ: Lawrence Erlbaum, 1984. 253-86.
- Paul, Danette. "Introducing Chaos into Scientific Discourse: A Study of Reception and Effect of Rhetorical Devices in Revolutionary Science." Diss. The Pennsylvania State University, 1996. Ann Arbor: UMI, 1996. DAI-A 57/04, p. 1602.
- _____, and Davida Charney. "Introducing Chaos into Science and Engineering: Effects of Rhetorical Strategies on Scientific Readers." *Written Communication* 12 (1995): 396-438.
- Pressley, Michael, and Peter Afflerbach. *Verbal Protocols of Reading*. Hillsdale, NJ: Lawrence Erlbaum, 1995.
- Rubin, Jeffrey. *Handbook of Usability Testing*. New York: Wiley, 1994.
- Schrivver, Karen. *Dynamics in Document Design: Creating Text for Readers*. New York: Wiley, 1997.
- Schumacher, Gary, and Robert Waller. "Testing Design Alternatives: A Comparison of Procedures." In *Designing Usable Texts*. Eds. Thomas Duffy and Robert Waller. Orlando, FL: Academic Press, 1985. 377-403.
- Slobodkin, L. B. "Foxes and Hedgehogs: A Look at Four Books by Celebrated Scientists." *American Scientist* 76 (September-October, 1988): 503-04.
- Smagorinsky, Peter, ed. *Speaking About Writing: Reflections on Research Methodology*. Sage Series in Written Communication, Vol. 8. Thousand Oaks, CA: Sage, 1994.
- Swales, John, and Hazem Najjar. "The Writing of Research Article Introductions." *Written Communication* 4 (1987): 175-91.
- van der Meij, Hans. "Catching the User in the Act." In *Quality of Technical Documentation*. Eds. Michael Steehouder, Carel Jansen, Pieter van der Poort, and Ron Verheijen. Amsterdam, The Netherlands/Atlanta, GA: Rodopi Press, 1994. 201-10.
- Winsor, Dorothy. "The Construction of Knowledge in Organizations: Asking the Right Questions about the Challenger." *Journal of Business and Technical Communication* 4 (1990): 7-20.
- _____. "Communication Failures Contributing to the Challenger Accident: An Example for Technical Communicators." *IEEE Transactions on Professional Communication* 31 (1988): 101-07.
- Wright, Patricia. "Quality or Usability? Quality Writing Provokes Quality Reading." In *Quality of Technical Documentation*. Eds.

Michael Steehouder, Carel Jansen, Pieter van der Poort, and Ron Verheijen. Amsterdam, The Netherlands/Atlanta, GA: Rodopi Press, 1994. 7-38.

_____. "Writing Technical Information. In *Review of Research in Education*, vol. 14. Ed. Ernst Rothkopf. Washington, DC: American Educational Research Association, 1987. 327-85.

Wyatt, David, Michael Pressley, Pamela El-Dinary, Shelly Stein, Peter Evans, and Rachel Brown. "Comprehension Strategies, Worth and Credibility Monitoring, and Evaluations: Cold and Hot Cognition When Experts Read Professional Articles That are Important to Them." *Learning and Individual Differences* 5 (1993): 49-72.