———. 2007. Mobility and composition: The architecture of coherence in non-places. *Technical Communication Quarterly* 16 (3): 279–309.
———. 2009. Recycled writing: Assembling actor networks from reusable content. *Journal of Business and Technical Communication* 24 (2): 127–163.
Wegner, Diana. 2004. The collaborative construction of a management report in a municipal community of practice: Text and context, genre and learning. *Journal of Business and Technical Communication* 18 (4): 411–451. doi:10.1177/1050651904266926.
Winsor, Dorothy A. 1996. *Writing like an engineer: A rhetorical education*. Mahwah, NJ: Erlbaum.
Witte, Stephen P. 1992. Context, text, intertext: Toward a constructivist semiotic of writing. *Written Communication* 9:237–308.
Wolfe, Joanna. 2002. Annotation technologies: A software and research review. *Computers and Composition* 19:471–497.
Yates, JoAnne, and Wanda Orlikowski. 2002. Genre systems: Structuring interaction through communicative norms. *Journal of Business Communication* 39 (1): 13–35.
Yin, Robert K. 2003. *Case study research: Design and methods*. Third edition. Thousand Oaks, CA: Sage.
Zachry, Mark. 2000. The ecology of an online education site in professional communication. In *Proceedings of IEEE professional communication society international professional communication conference and Proceedings of the 18th annual ACM international conference on computer documentation*, 433–442. New York: IEEE Educational Activities Department.
Zuboff, Shoshana. 1988. *In the age of the smart machine: The future of work and power*. New York: Basic Books.

BARBARA MIREL

12 How Can Technical Communicators Evaluate the Usability of Artifacts?

SUMMARY

Presenting information to audiences does not guarantee that they will value it or understand and use it effectively or efficiently. Before delivering information to audiences, it is crucial to evaluate whether presentations are truly useful and usable, and, if not, to determine how to improve them so that they are. Once artifacts are in audiences' hands, it is also important to evaluate whether evolving use signals a corresponding need to add improvements. Evaluating artifacts is one of the main contributions that technical communicators make in a project. Technical communicators evaluate information products for usability and usefulness, and based on findings, they recommend and help guide improvements. To evaluate digital information, various academic disciplines train usability specialists, but technical communicators are distinctively skilled in assuring that such artifacts succeed rhetorically with audiences. To conduct evaluations effectively, technical communicators need to understand and master the rhetorical skills, usability methods, and complex evaluation choices necessary for strategic assessments and usability reports. This chapter will help you develop these talents.

INTRODUCTION

"If you build it they will come." The creative spirit behind this expression of innovation can become a technical communicator's worst nightmare if the innovation is built with little regard for audience needs and purposes. Without adequately taking audience into account, innovations are not likely to be useful or usable. The following scenario illustrates such a case.

Collins is a technical communicator in a small firm that produces biomedical venture software. She aims to create effective user manuals based on initial documentation that developers draft to accompany their products. She also evaluates the usability of newly developed web applications through user performance testing. When testing applications, she

assesses the user manual and participants' uses of it as part of the tool evaluation.

The company in which Collins works develops life sciences applications for exploratory analysis. With them, scientists can explore huge volumes of diverse data and literature to uncover molecular-level relationships and processes that may influence mechanisms of a complex disease or other little understood physiological process. The applications let scientists query a database to retrieve information on genes, their functional traits, and their interactions. Retrieved information is displayed in networks of associated genes annotated by traits. The software provides other connected graphics for exploration as well, such as heatmaps (see figure 12.2). The visualizations are highly interactive and richly annotated with details such as functional attributes of genes, significance values for these attributes, biological pathways related to each gene, physical interactions between genes, and associated literature.

Collins learned today that she is to evaluate a new application and its developer-created manual. The developers of this application—BioConcept—believe the software and manual are both ready for user performance evaluation. The application aims to give scientists a quick means to search more than three million genes to uncover specific sets of potentially influential genes based on statistically significant functional traits. Collins is familiar with BioConcept from bimonthly meetings in her marketing group. Yet she has not been privy to the intricacies of design and development that went into creating this application. The developers' eagerness conveys that they expect BioConcept to effectively support users' needs and tasks even though it has not been subjected to either a user needs assessment or any iterative prototyping evaluation.

After an initial review of the software and manual, Collins sees that neither is ready for user performance testing. Neither meets common usability standards. If Collins tests these artifacts with users now, the tests will uncover only the problems already obvious to her. Collins decides on a different evaluation method known as heuristic evaluation. This assessment method involves applying a set of usability standards (also called heuristics) to interfaces and manuals. Usability heuristics have been established for a long time in the fields of technical communication and human-computer interactions. Collins has adapted these heuristics to the practices of exploratory analysis in the life sciences and validated them with usability colleagues and through pilot heuristic evaluations. She hopes that by running this usability inspection method, she can provide solid data to guide developers in improvements. After improvements, she hopes BioConcept and its manual will be ready for user performance testing.

At this point, Collins's situation highlights two choices for usability evaluations—inspections and user performance testing. As described later in this chapter, many more choices exist. A single chapter cannot explore usability in all its aspects for the various types of artifacts that technical communicators may assess. Consequently, the focus here is only on usability evaluation of software, web, and handheld applications. These industries are home to thousands of technical communicators like Collins, who may assess help systems, user documentation, tutorials, training materials, and user interfaces for websites, virtual realities, mobiles, games, tablets, software, and web applications. Additionally, in these evaluations, technical communicators implicitly have to assess aspects of a system's core functionality and architecture, because these aspects often contribute to usability and usefulness problems.

This chapter reviews usability research: it reveals the rhetorical nature of usability evaluation and provides findings from other studies on how to design and conduct your evaluations of digital artifacts. This review gives you details about evaluation methods such as usability inspection, field studies, formative and summative user performance testing, and mixed methods. It also covers success factors in writing usability reports. This background provides you with a good grounding for subsequently understanding three heuristic questions that should guide you in conducting evaluations:

1. What distinct skills do I have to apply to this usability situation?
2. What evaluation methods are best for the goals and circumstances?
3. What report choices will communicate convincingly and assure improvements?

To help you see how the questions apply to actual situations, we return to Collins's case at the end of the chapter.

LITERATURE REVIEW

In assessing artifacts, usability evaluators should assess both usefulness and usability. Usefulness is a value that users experience. It measures whether intended audiences find an artifact meaningful and valuable to their actual work flows and tasks as they want to do them. Usability, by contrast, is a property of the artifact. It measures and assesses whether the artifact's operations, displays, and content are easy to understand, use, access, learn, and navigate. Both usefulness and usability are critical

but—importantly—they are not mutually inclusive. For example, if an application does not enable domain specialists to work through "their inquiries according to situational and professional demands, even the easiest-to-use application is not useful" (Mirel 2004, 33).

Usability has many disciplinary homes besides technical communications. For example, usability evaluation is part of the training in information science, human-computer interaction programs, industrial design, performance technology, and the learning sciences. Like technical communications, these disciplines all offer courses related to usability at the undergraduate, master's, and doctoral levels. Yet technical communicators are distinctive in bringing rhetorical skills and training to usability evaluations (Cooke and Mings 2005). Rhetorical skills make evaluators sensitive to what makes language and the presentation of information work well for a particular purpose, audience, context, and medium. Language and information presentations can be the printed words or digitally represented communications flowing from an interface visually, symbolically, textually, and tactilely.

The research literature emphasizes the benefits of rhetoric as a distinct skill for usability evaluations. For example, Johnson, Salvo, and Zoetewey (2007) argue that usability without a rhetorical component is an "applied science" that misses many nuances of how people construct meaning from information in nonformulaic communication (interaction) situations. As an applied science, usability functions to formalize people's "human" approaches to digital knowledge work, so that these approaches can be turned into specifications for software requirements, procedural steps, and use cases. What this view omits is that information flowing from an interface or help system is a language establishing a discourse between a user and a tool (and implicitly the tool or text developer). As Sullivan (1989) notes, how audiences dynamically access, browse, interpret, understand, and act on information that flows from user interfaces or help systems is mediated through language. The language communicates to users the affordances a tool offers—and does not offer—for their needs and purposes. Whether visual, verbal, or symbolic, the languagelike representations communicating to users at the interface have rhetorical dimensions. That is, the effect of a system or document design depends on whether it connects with users' actual purposes, contexts, domains, roles, reasoning, constraints, and conventions.

Consequently, assessing how interfaces or help systems are received, taken up, applied and valued by intended audiences requires rhetorical expertise. Anscheutz and Rosenbaum (2002) relate through several case histories how rhetorical skills have helped many technical communicators expand their workplace roles and responsibilities. In these cases, individuals take on lead roles in usability testing, user-centered system design, quality assurance, and user-interface design and development.

In addition to rhetorical skills, usability evaluators need to have a tool kit of methodologies from which to design and carry out evaluation projects for various situations. A large body of research, best practices, and guidelines exists that details diverse methodologies and protocols—both qualitative and quantitative, formative and summative (Rubin and Chisnell 2008; Krug 2005; Barnum 2002; Blakeslee and Fleisher 2007; Dumas and Redish 1993; Redish et al. 2002; Blandford et al. 2008; Sutcliffe et al. 2000; Cockton and Woolrych 2001; De Jong and Shellens 2000; Rosenbaum 2008; Ummelen 1997; Spyridakis et al. 2005; Kushiniruk and Patel 2004). Familiarity with this literature is important to guide the formulation of evaluation goals and methods and to assure that they fit a given multifactored usability situation. The implications of various situational factors for goals and methods also are discussed in the research literature (e.g., Barnum 2002; Dumas and Redish 1993). For example, the stage of artifact development is an important determinant of optimal methods for evaluation, as are the types and complexity of tasks afforded by the artifact and the diversity of its target users.

Ideally, usability evaluations are performed after many other user-centered activities for artifact design and development have already taken place. For example, context-based audience and needs assessments should occur and influence design before usability testing. Additionally, project management structures should be in place to assure that needs assessments and their influence on prototype designs are built into the front end of the development cycle from the start. These activities are addressed by other chapters in this book. If these activities do not occur before a usability evaluation, the evaluation is likely to uncover deep conceptual, architectural, and communication problems, at a point when it is harder to remedy them and can add significant costs and time to production. As Scotch, Parmanto, and Monaco (2007) show, if user performance testing demonstrates inadequate baseline usability in an artifact due to an absence of early assessments, users cannot conduct the full range of tasks required for evaluation. The evaluation, consequently, will not be able to uncover deeper problems that are likely to be detrimental to users' task flows in their actual analyses.

As part of their tool kit of methods, usability specialists can consider conducting various modes of usability inspection (Nielsen 1993; Hollingsed and Novick 2007)—cognitive walk-throughs, expert reviews, and heuristic evaluations. In cognitive walk-throughs, a usability specialist

or domain expert performs predefined tasks that cover the types of goal-based interactions and reasoning that developers intend the application to support in actual use. Outcomes show actions afforded by the interface that are prone to error or other difficulties. In expert reviews, usability specialists, domain experts, or both intuitively evaluate whether artifacts reflect acceptable quality, based on their respective professional expertise. Heuristic evaluations, mentioned earlier in Collins's case, apply severity rankings to judgments about satisfying the specified standards (heuristics). Severity rankings might include these levels:

- Level 5: a catastrophic error causing irrevocable loss of data or damage. The problem could result in large-scale failures that prevent people from doing their work. Performance is so bad that the system cannot accomplish business goals.
- Level 4: a severe problem causing possible loss of data. A user has no work-around, and performance is so poor that the system is universally regarded as "pitiful."
- Level 3: a moderate problem causing wasted time but no permanent loss of data. There is a work-around to the problem. Internal inconsistencies result in increased learning or error rates. An important function or feature does not work as expected.
- Level 2: a minor but irritating problem. Generally, it slows users down slightly, involving poor appearance or perceptions, and mistakes that are recoverable.
- Level 1: a minimal error. The problem is rare or is tied only to minor cosmetic or consistency issues (Wilson and Coyne 2001).

Heuristic evaluation standards are often generic, and better outcomes occur when standards are adapted to the domain targeted by an application (Cockton et al. 2007; Mirel and Wright 2009; Cockton and Woolrych 2001). In terms of all inspection methods, research shows that no single method can achieve adequate usability (Nielsen 1994). Rather, these methods need to be combined with other evaluation approaches, such as user performance testing, interviews and surveys, and usage-log analysis.

Evaluations through user performance testing are either formative or summative. Formative testing occurs early and often, and it generates findings that can progressively improve an artifact during development. In formative testing, qualitative methods are often used because, at this early stage, too little is known about factors contributing to usefulness for a given class of tasks and work flows to have valid constructs available for quantitative analysis. Formative scenario-based testing and qualitative analysis help define such constructs and standards of excellence. These

constructs and metrics often become the criteria by which an artifact is assessed in subsequent summative testing.

Summative testing occurs at the end of development and validates an application. Summative evaluations determine if an application and its documentation effectively and efficiently address audience, purpose, context, and tasks and if they meet standard usability criteria. This testing is typically quantitative. This chapter does not describe summative performance testing: the precision with which its quantitative methods need to be applied are covered in other sources, for example, Sauro and Lewis 2009; Evans, Wei, and Spyridakis 2004; Kirakowski 2005; Saraiya, North, and Duca 2005; Gray and Salzman 1998; and Hughes 1999.

Whether an evaluation is formative or summative, the evaluator needs to determine the appropriate unit of analysis for the user performance test. The unit of analysis specifies the focus. For example, the focus might be on individual or collaborative performance, or on user performance on predefined tasks or user-defined tasks. In evaluations of early yet fully functional prototypes, evaluators often run user performance tests in the field to identify contextually embedded demands for usefulness. Evaluators gather data as users perform their actual software-supported work in naturalistic work settings (Mirel 2004). In such field studies, the unit of analysis may be individual or collaborative performance of just certain tasks or performance of any on- and offline tasks having to do with a targeted problem. Field studies with the latter unit of analysis—any on- and offline tasks related to the problem—are especially effective for evaluating usefulness. This unit of analysis requires comprehensive data collection. It involves collecting observation and interview data. It also involves gathering and analyzing log data (automatically generated records of users' interactions with a program/website when they are not being observed); diaries; and other self-reports (Jarrett et al. 2009; Ivory and Hearst 2001; Spyridakis et al. 2005).

A usability evaluation method that often comes after field studies focuses on testing user performance on just certain specified features of a prototype (Snyder 2003). The unit of analysis is typically an individual's use of the specified features for predefined tasks. In this testing, evaluators often work hand in hand with developers to run quick cycles of prototype development, usability testing, prototype revision, and retesting. This approach is called rapid iterative testing and evaluation (RITE). It looks at program-defined, low-level actions, often actions that apply to higher-level tasks and work flows that users revealed during earlier field testing. Evaluators test just a handful of users (five or so), observing and gathering data on them as they interact with the prototype. Data may in-

clude time on tasks, errors, success rates, and recovery times from errors as well as satisfaction feedback (Medlock et al. 2005). This testing method is optimal when team members are experts in usability and when the targeted user tasks are well-structured. In these cases, usability experts are able to translate findings into redesigns and improvements quickly and to collaboratively set priorities and implementation choices with developers and other stakeholders (Rosenbaum 2008).

Because of organizational circumstances, as in Collins's case, usability specialists in an organization may not be able to conduct early field studies and iterative prototype testing. Organizations vary in their awareness of the importance of incorporating usability processes into the product-development cycle from inception on. In cases in which no early field tests or RITE testing has been performed, usability performance tests take on the burden of having to assess both usefulness and usability. To combine the evaluation of usefulness and usability, for example, predefined tasks and scenarios in the user performance test need to generate (1) findings on usability issues of concern (e.g., smoothness of flow of interactions with features and functionality, error management) and (2) findings on fitness to purpose (adequate task support for the actual reasoning and domain knowledge that users apply in their work). As cited previously, a good deal of research exists to guide informed decisions about test design and methodology for user performance testing in controlled environments. Research in the literature also addresses various issues specific to certain media (e.g., mobile, virtual reality, speech recognition), cross-cultural systems, degree of task complexity, domains (e.g., health, finance), demographic segments, and accessibility.

In field studies and controlled settings, user performance evaluations need to include appropriate sampling methods and sample sizes (Koerber and McMichael 2008). Strict criteria exist for quantitative studies, and it can be helpful to consult with statistical experts in designing these studies. For formative and largely qualitative evaluations, samples of users can be recruited through convenience sampling or purposive sampling methods. Convenience sampling, that is, selecting users based on availability alone, is not sufficient when users need to bring specific knowledge and experience levels to the tasks or scenario. A better method for this situation is purposive sampling, which involves selecting users based on set criteria for user traits. Typically, twenty users is a good sample size and acceptable for qualitative evaluation (Miles and Huberman 1994).

Conventional methods for running user performance evaluations include having at least two usability evaluators observe user performance sessions, upon consent from users. The evaluators ask users to think out

loud as they conduct predefined tasks or scenarios. That is, they ask users to articulate thoughts such as intentions, choice points, and reasoning as they proceed through their work. The usability evaluators also train users briefly in think-aloud processes before testing (Boren and Ramey 2000). During the task performance sessions, software interactions and think-alouds often are video- and audiorecorded by screen-capture software. Evaluators also decide whether to set time limits on the performance of a task, based on whether such constraints are consonant with their evaluation goal. Evaluators observe each user session without intervening, and take notes to guide later analysis of the raw data. At the end of each user session, they often ask users to fill out standard satisfaction surveys (Brooke 1996; Kirakowski 1996; Kirakowski and Cierlik 1998). Satisfaction surveys—their reliability, metrics, significance of outcomes—have their own art and science, and usability evaluators should be familiar with the literature about them (Sauro and Kirklund 2005; Lewis 2006; Sauro, and Lewis 2009; Bangor, Kortum, and Miller 2008). Evaluators debrief after each session, sharing their perceptions and highlights.

To analyze the formative user performance and satisfaction data, evaluators typically follow standard qualitative methods. They characterize patterns, exceptions, and themes in users' performance behaviors, knowledge, and affective reactions (Krippendorff 2004; Barton 2002; Creswell and Clark 2006; Blakeslee and Fleisher 2007). To start, they holistically view the video and audio recordings of user sessions several times. Then, for each user session, they tag various uses of artifact functions and features and user behaviors for important traits and indicators of performance efficiencies and effectiveness. These may include, for example, types of information-seeking behaviors (e.g., access, monitor, search, browse, extract, chain, analyze; Makri, Blandford, and Cox 2008); demonstrated program-related problems, impasses, and errors; error-recovery instances; categories of interactivity (e.g., selection, filtering, navigating within a screen, navigating across screens, backtracking); elapsed time on certain tasks; and task boundaries. Screen-capture analysis systems and content analysis tools can expedite these analyses.

Evaluators also transcribe the think-alouds and interview responses, and analyze them to abstract patterns shared across user cases. Additionally, they extract exceptions and themes characterizing expressed and acquired knowledge, modes of reasoning, affective reactions, and patterns and exceptions during performance and post hoc interviews. Unlike qualitative methodologies that aim primarily to build grounded theory, these qualitative analyses may have theory-building outcomes, but they also give high priority to action. This action includes theoretically grounded

rationales, recommendations, and user-oriented specifications for enacting design modifications and improvements.

As the research literature suggests, large amounts of data are gathered and analyzed to generate assessments and to recommend improvements. It is no small feat to turn findings from evaluations into communications and recommendations that prompt developers and other stakeholders to set the right priorities for improvements and to construct truly effective improvements. Evaluation specialists agree that writing high-quality reports "with recommendations that are taken seriously by the product team" is a core aspect of usability testing (Brady 2004, 67). Unfortunately, this aspect of an evaluator's role is underresearched. Few studies cite results about the best ways to compose effective formative evaluation reports and recommendations to direct evolving improvements and priorities for greater usefulness and usability.

Standards and guidelines for writing effective summative reports are better established and disseminated than standards and guidelines for formative evaluation reports (Industry Usability Reporting Project 2001). A well-delineated report format for summative testing, supplemented with examples, can be found at the National Institute of Standards and Technology (NIST) website (http://zing.ncsl.nist.gov/cifter/TheCD/Cif/Readme.html).

Despite a relative paucity of research into formative evaluation reporting, investigators have gained some important insights into effective strategies for structuring reports. For example, a qualitative analysis of formative reports and recommendations composed by seventeen teams of experienced usability professionals shows fifteen organizing patterns characterizing these reports' section structures. In general, the researchers argue that some or all of the elements in table 12.1 should be included in reports (Theofanos and Quesenbery 2005).

As an overarching composition principle, researchers highlight the need for rhetorical effectiveness—something that should be second nature to technical communicators (Theofanos and Quesenbery 2005). For example, research stresses that writing choices should be shaped by the business context, its conventions, constraints, and priorities. Rhetorically, choices in reporting also should be shaped by the writer's relationships with the intended primary and secondary readers, their prior knowledge, their likely assumptions and misconceptions, and the questions that the evaluation tried to answer. Finally, rhetorical choices in reporting depend on the current phase of the development cycle, the type of artifact being evaluated, and the buy-in likely from the audiences.

Table 12.1. *Elements to include in evaluation reports*

1. Title page	6. Overall test environment	11. Metrics
2. Executive summary	7. Participants	12. Quotes, and screen shots
3. Teaching about usability	8. Tasks and scenarios	13. Conclusions
4. Business and test goals	9. Results and recommendations	14. Next steps
5. Method and methodology	10. Recommendation details	15. Appendices

Researchers also suggest content strategies. These strategies include presenting recommended improvements in the form of screen mock-ups with call-outs, including screen shots that depict the problem, and providing quotations from users as a means to bring the audience "in touch with users . . . and building awareness of user needs" (Theofanos and Quesenbery 2005, 34). Another means for eliciting a positive response from audiences composed of developers or managers is to include usability successes as well as problems. Researchers find that it is also important to categorize problems in the report by the user experience issues to which they relate, not by the program feature. For well-structured work, many such categorical schemes are available in the research literature (Fu, Salvendy, and Turley 2002). For complex work, Blandford et al. (2008) and Sutcliffe et al. (2000) provide a number of problem categories and descriptions that achieve this effect well with diverse readers—developers, managers, and users (detailed in a later section). In addition, problems and associated recommendations should have severity rankings and provide clearly written criteria for each rank.

For all these issues covered by the research literature—the necessary skills for evaluation, the application of the right methods at the right time, and the development of persuasive reports—a set of heuristic strategies can help you systematically bring them into your actual usability evaluations. The next section looks at these heuristics.

HEURISTIC

A good deal of work goes into preparing for, conducting, and reporting on usability evaluations. Organizing this work by three main heuristic questions will help direct your efforts.

WHAT DISTINCT VALUE AND SKILLS DO I HAVE TO APPLY TO THIS USABILITY SITUATION?

Your rhetorical expertise will be relevant to every usability situation. However, each usability situation calls for different combinations of other skills and knowledge and emphases. Evaluating artifacts effectively de-

pends on the right combination. Among the skills you will need to variously combine with rhetorical expertise are

- knowledge and skills of universal usability (accessibility for different disabilities);
- knowledge of niche-based communication media, for example, visualizations;
- awareness of standard resources and setups for usability laboratories;
- awareness of software packages/technologies that can facilitate data collection and analysis (e.g., software for screen capture, usage-log analysis, content analysis);
- knowledge and skills in the subject matter and domain in which you work;
- a working knowledge of human cognition and reasoning, from novice through expert, including cognitive psychology, sociocognition, social interactionism and constructivism, distributed cognition, and actor network theory;
- an understanding of the technical logic and technological efforts required for development in areas related to providing various support to users; and
- a working knowledge of the relationship between user needs and goals, on the one hand, and design and development choices for user interfaces, manuals, and software design, on the other hand.

WHAT EVALUATION METHODS/TEST DESIGNS ARE BEST FOR THE GOALS AND CIRCUMSTANCES?

All the questions that follow direct you toward matching methods and evaluation designs to specific goals and conditions of a usability situation.

- *Why evaluate?* You should state and convincingly support why a usability evaluation is in order.
- *What is expected?* You should find out what developers think needs to be evaluated and fixed—and compare it to what you think needs to be fixed. Ideally, these will be in sync; practically, they often are not (Howard 2008). Find out, as well, the assumptions and other obstacles that may lead stakeholders to misconceive the meanings and purposes of usability as well as their perceptions about evaluations that are feasible and problems that can be uncovered through specific evaluation methods.
- *What constrains the evaluation?* You should be clear about the resources, time, effort, and expertise that are available for usability evaluations, all of which constrain the choices of evaluation goals and methods. Also, identify how much and what is already known about usability issues relevant to this application and what evaluations have produced this knowledge. For example, what needs assessments have been completed to guide application development thus far, and who conducted them? What are the quality and substance of the user-oriented results of the needs assessments? What development processes are followed, and where do usability assessments figure into the development processes?
- *Who are the users?* It is important to make it clear to yourself and stakeholders who you intend to define as the targeted primary, secondary, and tertiary audiences of the application and manual. Be clear how the target audiences affect the scope of your evaluation. It is not uncommon for audience definitions and their ranked importance to be unclear. Therefore, explicitly negotiate with developers and other stakeholders to reach a consensus about target audiences before the evaluation. For example, biostatisticians may experience tasks with the program differently from laboratory scientists.
- *How will the artifacts help users?* You need to construct evaluation goals and methods attuned to the goals of each audience—their tasks and the support and enhancements the artifact should provide for them. Set your scope wide enough to include tasks for which users will want to control interactivity.

WHAT REPORT CHOICES WILL COMMUNICATE CONVINCINGLY AND ENSURE IMPROVEMENTS?

The research literature on report writing discussed earlier suggests important issues to consider. Making choices about high-level section headings, for example, provides a good first cut in determining the scope and content of the report and necessary connections with primary and secondary audiences. Good candidates for high-level headings are detailed in the literature review.

High-level section headings alone, however, will not give your report a user-experience perspective. That perspective depends as well on your choices of content, emphasis, and subheadings within many of the major headings (e.g., business and test goals, screen shots and video). Of all the major headings, arguably the most important one for orienting readers to a user point of view is the results-and-recommendation section. In this section, certain subheadings are better than others for evoking a user experience perspective on reported problems and recommended im-

provements. For example, traditional subheadings that focus on program operations, functions, and features conjure a system focus rather than a user focus for readers. User-centered names for subsections vary, with many examples provided by Sutcliffe et al. (2008) as well as Blandford et al. (2000), as follows:

- Missing functionality for conceptual reasoning and user tasks
- Inadequate or partial functions that fall short of user needs and expectations
- Viscous support, that is, too many actions, high costs for small moves, difficulty in specifying the action sequence pertaining to a domain-based task
- Visual attention not matched to user needs, that is, inability to detect what needs to be detected, defaults that do not draw the eye to selectively important items or relationships for a task
- Clarity of "What do I see, and what can I do with it?"
- Clarity of "What did that do?" feedback
- Clarity of "How do I get to what I want to do?"
- Clarity of "Where have I been, and what do I know?"
- Imprecision in seeing and/or doing, that is, difficulty carrying out actions or discriminating

The examples above highlight the context of user experiences that you can further narrate, diagram, or otherwise represent within each subsection.

Why should you pay such close attention to your subsection headings? In part, doing so will help avoid the common phenomenon of developers glossing over such overused subsection headings as "ease of use" and "ease of access" and looking straightaway only for implicated program features. These readers miss the coherence and completeness that users need to experience to interact with an artifact usably and usefully, and subsequent development efforts may repeat the same ill-conceived design choices in other artifacts.

To show these three heuristic questions in real situations of usability evaluation, we return to the case of Collins and the usability evaluation of BioConcept.

EXTENDED CASE: A RETURN TO EVALUATING BIOCONCEPT

In Collins's case, usability has not been addressed until the end of the development cycle. Had the development team incorporated ongoing usability assessments into the development cycle from the start, BioConcept most likely would have been ready for formative user performance tests, as its developers hoped. Then Collins could have followed the testing methods described in the literature review. So far, the lessons learned from Collins's case include the following:

- If project management does not include early needs assessments and iterative testing, the evaluation situation is likely to be less than ideal.
- Heuristic evaluations and other quick inspection methods are a start, but they are not sufficient in themselves.
- In software and manuals, giving access to huge volumes of information is not sufficient for users' analytical purposes. Presentations must accord with users' *flows* of analysis. The right content in the right verbal and visual forms needs to be presented with effective explanations; and the right level of user-controlled interactivity must be included and communicated. These are, in part, rhetorical choices.

As the case so far shows, Collins needed communication expertise in this situation—as addressed by the first heuristic question. She saw scientists' potential interactions with BioConcept as a human-computer *communication* act. Specifically, Collins was concerned that the initial software lacked—and therefore did not adequately communicate—the full range of content and interactivity relevant to scientists' analytical purposes. In critically reviewing the software and the manual, she did not simply make sure it "contained" rich information and innovative features, regardless of how well their design fit users' purposes and context. Rather, she took a perspective centered on users' holistic interactive experiences with the artifact. Based on research findings and her own experiences, she believed that when evaluations focus only on separate components of this experience instead of the whole, they end up assessing effects of just those formalizable aspects of users' experiences, for example, the low-level actions of selecting a data item or clicking an interface widget for an operational step. It is important for users to conduct these operations easily, but ease of use is moot if they cannot do the tasks and functions that meet their analytical objectives in the first place.

Collins initially chose heuristic evaluations to identify obvious problems that needed fixes. Following standard practices of having several raters, Collins brought in three colleagues with whom she had conducted heuristic evaluations on similar tools. Table 12.2 shows an excerpt from their heuristic evaluation of BioConcept.

To extend the case further, after the developers revised the application and manual based on results of the heuristic evaluation, the artifacts still lacked many of the capabilities and interface designs that would match

Table 12.2. *Sample problems found in the heuristic evaluation*

Heuristic (usability standard)	Severity rating (0 = no problem; 5 = major problem)	Comments about usability
Does it clearly show if there are no query results?	4	Shows a 0 but should have a sentence like "No results found for 'csfir'" or something similar.
Is it easy to reformulate the query if necessary?	5	Big problem: if you go back to the search screen from the explorer, you lose your query and the search results.
Are the results transparent as to what results are being shown and how to interpret them?	4	Could use a header that indicates what we're looking at; e.g., "45 biomedical concepts found matching 'cancer.'"
Is there ability to undo, redo, or go back to previous results (e.g., history tracking)?	5	No history tracking.
Are the mechanisms for interactivity clear?	4	Could use more labeling or tool tips.
Can you access the necessary data to assure the validity of results (i.e., the sources of the results)?	4	Can't get to sources (e.g., can't click on MeSH term and get to MeSH).
Can results be saved?	5	No saving option.

users' purposes and practices. For example, scientists would expect to compare many conceptual groupings to uncover overlapping genes among groups. Yet BioConcept did not give them an easy and efficient way to make comparisons.

The manual also had problems (see figure 12.1). It did not tie program operations to analysis-based task objectives. Nor did it explain the tool in ways that would connect and resonate with scientists' exploratory intentions and standards of practice and validation. Despite revisions, explanations were scanty in the manual, and those that were included were copied verbatim from an article that the developers wrote for an audience of computer scientists, biostatisticians, and computational biologists, not

Gene-Set Relation Mapping – Heatmap View [Explanation]

The values used in creating the heatmap are defined by the counts of the enrichment concept pairs that a gene belongs to, and the genes and concepts are clustered using complete linkage hierarchical clustering with the euclidean distance measure. Color of columns ranges from black (gene belongs to no enriched concepts) to bright red (gene belongs to the most enriched concepts).

BioConcept provides a heatmap view of your gene set and its enriched concepts. To view relationships by using a heatmap view: [Procedure]

1. In the Concept Explorer window, select the concepts that you would like to view. You may click the **Select All** link in the chart area if you would like quickly to select all of the enriched concepts.
2. Click the **Draw Heatmap** button on the bottom of the Concept Explorer window.

Heatmap Viewer

Heatmap Characteristics	Meaning
X-axis	Your gene list
Y-axis	Enriched concepts
Color	Redness indicates number of enriched concepts that include a specific gene.
Clicking *Draw Network Graph* button	Displays the Network Graph view.
Clicking *Network View* button	Displays the Network Graph view. When in the Network Graph view, this button changes to Heatmap view, which allows you to switch between both views.
Clicking *Explorer* button	Brings up Concept Explorer screen (click the X in the upper right-hand corner of the Concept Explorer screen to return to the Network Graph View).
Clicking *Export Graph* button	Creates a PNG file image of the screen in a separate window.

Figure 12.1. Explanations and procedures in the developer-composed documentation. This passage is taken verbatim from an article that is not aimed at the primary user audience. Procedures lack task-driven goals and outcomes as framing devices.

life science researchers (figure 12.1). Similarly, screen shots did not help users interpret displays in relation to a sample analytical task (figure 12.2).

Collins recognized these problems without needing to put the application and manual in front of users. Some problems are likely to require new functionality and interface modifications and will not be quick fixes. Collins faces a decision on usability evaluation at this point. Should she conduct a user performance test or push for further improvements first? Factors affecting Collins's decision about the next steps to take are tied to the questions presented above, on evaluation methods and test design.

Collins knows that any decision is going to involve trade-offs. On one hand, she could conduct user performance testing and focus on only the narrow set of discrete features and operations that are ready with baseline usability and usefulness for scientists' tasks. This choice would keep the development cycle on schedule. But it would let some major problems slide, such as insufficient support for making comparisons. On the other hand, she could push for more improvements before conducting a user performance test. This choice would require investing more time and effort than budgeted in the development plan. She would need to make a persuasive case to management for this greater investment. In presenting her case, she could discuss the likelihood of mismatches between BioConcept and scientists' complex analytical needs for discovery, which would likely frustrate early adopters and make them skeptical about the quality of the firm's software. These early adopters are vital to the company for early user feedback after deployment, and they might abandon their loyalty to the software.

Again communication skills are a necessity, now for persuasion. Better outcomes in product design and development require a shared understanding across stakeholders of why certain evaluations are being conducted at a given point, based on application quality and progress.

Collins decides to pursue the second option. This option will push back the production schedule—but do so at an early enough point to make such adjustments without prohibitive costs or revenue losses. Collins's persuasive memo to management (copied to developers) is different from the usability test reports addressed by the third heuristic question. The persuasive memo, nonetheless, serves similar goals by providing empirical support for the usability judgments that Collins makes. For this support, Collins may informally put the current versions of BioConcept and its manual in front of a "friendly" early adopter of the firm's software. A trial user performance can identify successes, flaws, and limits to cite in the memo that go beyond Collins's subjective expertise.

Ideally, management will concur, the tool and manual will be improved,

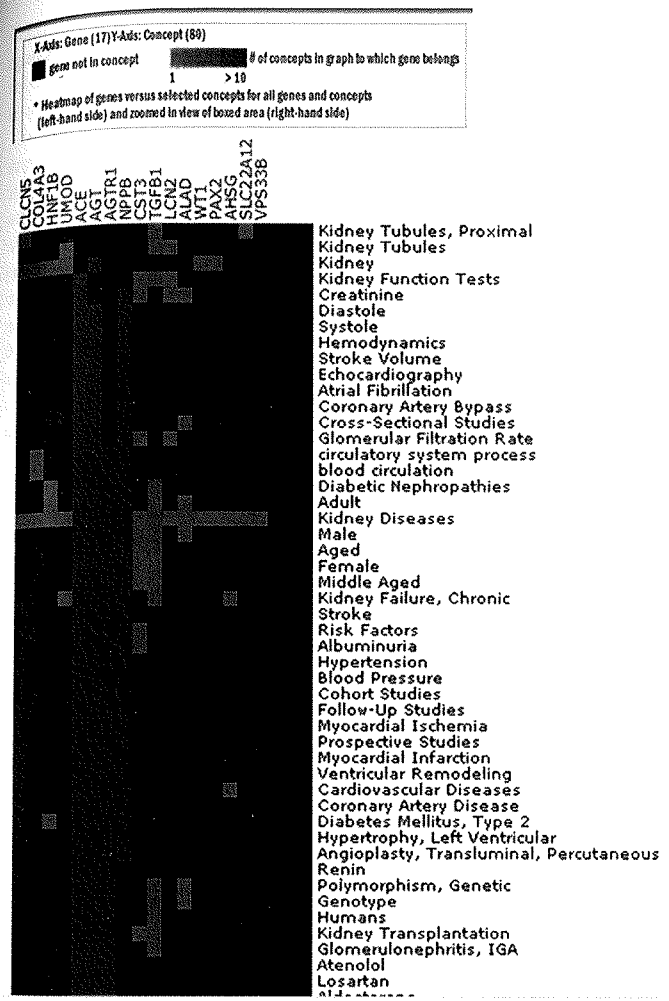

Figure 12.2. Heatmap view. The screen shot of the interactive visual presented in the documentation lacks labels and other explanations and cues to facilitate interpretations and applied meanings to an analysis objective.

and Collins will conduct the user performance evaluation. She will then report results in ways that effectively "talk" to stakeholders, addressing the issues implicit in the third heuristic.

At that time, she can present usability successes and decide on the structure of the problem sections. Because of her rhetorical orientation, she will organize sections by users' cognitive tasks, that is, by users' demonstrated task-based reasoning and behaviors. For the specific domain-based analysis conducted with BioConcept, the usability report subheadings might highlight users' comprehension of terms; their dependence on cues for staying oriented in analysis; their understanding of relevant molecular relationships based on specific workspace features, functions, and content; their validation of relationships; and the data manipulations they need and expect to do for specific types of task-based goals and reasoning. In using these higher-order cognitive tasks to structure the report, Collins will remind readers that choices about design modifications directly impact users' performance and cognition. She can subdivide these cognitive task categories into types of usability and usefulness problems and discuss tool features and functions implicated in each problem (Molich, Jeffries, and Dumas 2007). She will be guided by the categories synthesized from Sutcliffe et al. (2000) and Blandford et al. (2008) mentioned earlier in the heuristic related to writing reports.

Collins's report also includes severity rankings, her recommended priorities for fixes, and suggested designs for fixes. Application problems have behavioral consequences for users' practices and reasoning, and readers must be aware of these consequences. She aims for recommendations and suggestions with enough detail to guide the actual implementation of modifications. Importantly, she recommends fixes to problems that cause immediate usability shortcomings, but she also looks at the big picture. Her report will project the consequences of recommended fixes and enhancements. When stakeholders and Collins meet to review her report, they will work as a team to negotiate priorities in improvements. Her severity rankings and criteria will contribute to ultimate decisions about priorities for improvements.

Combined, these various structures and content will add impact to her report. But evaluation reports in themselves cannot carry the weight of the usability influence that Collins and others like her need to assert in their organizations. This influence is ongoing, interpersonal, and instructive. As Collins's case shows, her report on the user performance testing will be one part of a larger flow of communications about the quality and usability of BioConcept. From her earlier heuristic evaluation report through her persuasive report to management, she has been involved in

negotiating with diverse stakeholders about improving BioConcept with convincing evidence. Successful collaborations and respect for usability roles in a product team are prerequisites for achieving the goals of usability evaluation reports.

CONCLUSION

This window into evaluation situations that you might encounter when assessing software, web applications, and/or documentation reveals that strategic choices run throughout your processes of setting evaluation goals, determining methods, and reporting. Strategically, it is important to actively assume the role of usability evaluator and justify your value and suitability for the role to the product team. Additionally, knowing methods and having a tool kit of skills are necessary but not sufficient. It is vital to use them advantageously. You need to devise goals for evaluation that accord with the situation at hand and select—and possibly mix—methods. Strategic approaches to reporting are also important, including your choices for framing the content, structure, and media of communications to stakeholders in ways that maximize the chances that top-priority problems and enhancements will be acted upon effectively. Throughout this evaluation work, you will be establishing your role and identity in the organization as an irreplaceable usability expert whose skills and knowledge in technical communications and related areas add value to the development and dissemination of products.

DISCUSSION QUESTIONS

1. What are some important criteria for judging usefulness? What are some important criteria for judging usability? Which criteria for usefulness and usability overlap with each other, and which are distinct from each other?
2. Reflect on this claim by answering the questions following the claim: "Knowing how to evaluate an artifact so that it achieves the purpose of having audiences go along and act effectively on it is a communication and rhetorical art." What issues related to usefulness and usability, respectively, are rightfully communication issues and why? What issues related to usefulness and usability are not communication issues and why?
3. Go to your university's library website, and use its catalog-search page to explore a topic that interests you. Evaluate the set of pages resulting from the search, using the extract from the heuristic evaluation instrument included in this chapter. Include comments and severity rankings. Compare your ratings and comments with three or four of

your classmates. On what do you agree and disagree? What criteria were you each using to determine "level of severity"? With these three or four classmates, try to agree on criteria for defining severity level.

4. Which of the usability methods mentioned in this chapter interest you most? What else would you like to find out about them? Write at least three questions.
5. For what aspects of the usability evaluation situations and approaches described in this chapter do you feel most prepared? For what aspects do you feel least prepared? Explain in detail your perception of your preparedness.
6. Examine an application or game that you like but that, at times, frustrates you. For the frustrating portion, write a real-world task that participants in a user performance test could perform. Craft the task so that it will enable evaluators to gather data on how easy and useful this "frustrating aspect" of the technology is (or is not) for this particular task. Write a brief rationale for your task, explaining what data it might generate that could provide evidence of the tool's shortcomings and strengths.
7. If you were writing a report for only the product team on the user performance testing that Collins ultimately will conduct, what elements and content would you include and why? If you were writing for only the director and marketing group, what elements and content would you include and why? If you were writing to everyone, what elements and content would you include and why?

WORKS CITED

Anscheutz, Lori, and Stephanie Rosenbaum. 2002. Expanding Roles for Technical Communicators. In *Reshaping Technical Communication*, ed. Barbara Mirel and Rachel Spilka, 149–164. Mahwah, NJ: Elsevier.

Bangor, Aaron, Philip Kortum, and James Miller. 2008. An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction* 24:574–594.

Barnum, Carol. 2002. *Usability Testing and Research*. New York: Longman.

Barton, Ellen. 2002. Inductive Discourse Analysis: Discovering Rich Features. In *Discourse Studies in Composition*, ed. Ellen Barton and Gail Stygall, 19–41. Cresskill, NJ: Hampton Press.

Blakeslee, Ann, and Cathy Fleisher. 2007. *Becoming a Writing Researcher*. Mahwah, NJ: Erlbaum.

Blandford, Ann, Thomas R. G. Green, Dominic Furniss, and Stephann Makri. 2008. Evaluating System Utility and Conceptual Fit Using CASSM. *International Journal of Human-Computer Studies* 66:393–409.

Boren, Ted, and Judith Ramey. 2000. Thinking Aloud: Reconciling Theory and Practice. *IEEE Transactions on Professional Communication* 43:261–277.

Brady, Ann. 2004. Rhetorical Research: Toward a User-Centered Approach. *Rhetoric Review* 23:57–74.

Brooke, John. 1996. SUS: A Quick and Dirty Usability Scale. In *Usability Evaluation in Industry*, ed. Patrick W. Jordan, Bruce Thomas, Bernard A. Weerdmeester, and Ian L. McClelland, 89–194. London: Taylor and Francis.

Cockton, Gilbert, and Alan Woolrych. 2001. Understanding Inspection Methods: Lessons from an Assessment of Heuristic Evaluation. In *People and Computers XV*, ed. Ann Blandford and Jean Vanderdonckt, 171–192. Berlin: Springer.

Cockton, Gilbert, Alan Woolrych, and Darryn Lavery. 2007. Inspection-Based Evaluations. In *Human-Computer Interaction Handbook*, ed. Andrew Sears and Julie Jacko, 1172–1191. Boca Raton, FL: CRC Press.

Cooke, Lynne, and Sue Mings. 2005. Connecting Usability Education and Research with Industry Needs and Practices. *IEEE Transactions on Professional Communication* 48:296–312.

Creswell, John, and Vicki Plano Clark. 2006. *Designing and Conducting Mixed Methods Research*. Thousand Oaks, CA: Sage.

De Jong, Menno, and Peter Jan Shellens. 2000. Toward a Document Evaluation Methodology: What Does Research Tell Us about the Validity and Reliability of Evaluation Methods? *IEEE Transactions on Professional Communication* 43:242–260.

Dumas, Joseph S., and Janice Redish. 1993. *A Practical Guide to Usability Testing*. Norwood, NJ: Ablex.

Evans, Mary, Carolyn Wei, and Jan Spyridakis. 2004. Using Statistical Power Analysis to Tune-up a Research Experiment: A Case Study. In *Proceedings of IEEE International Professional Communication Conference*, 14–18. New York: IEEE Press.

Fu, Limin, Gavriel Salvendy, and Lon Turley. 2002. Effectiveness of User Testing and Heuristic Evaluation as a Function of Performance Classification. *Behaviour and Information Technology* 21:137–143.

Gray, Wayne D., and Marilyn Salzman. 1998. Damaged Merchandise? A Review of Experiments That Compare Usability Evaluation Methods. *Human-Computer Interaction* 13:203–261.

Hollingsed, Tasha, and David Novick. 2007. Usability Inspection Methods after 15 Years of Research and Practice. In *Proceedings of the 25th Annual ACM International Conference on Design of Communication (SIGDOC '07)*, 249–255. New York: ACM.

Howard, Tharon. 2008. Unexpected Complexity in a Traditional Usability Study. *Journal of Usability Studies* 3:189–205.

Hughes, Michael. 1999. Rigor in Usability Testing. *Technical Communication* 46:488–494.

Industry Usability Reporting Project. 2001. Common Industry Format for Usability Test Reports. ANSI/INCITS 354-2001. New York: American National Standards Institute.

Ivory, Melody, and Marti Hearst. 2001. The State of the Art in Automating Usability Evaluation of User Interfaces. *ACM Computing Surveys* 33:470–516.

Jarrett, Caroline, Whitney Quesenbery, Ian Roddis, Sarah Allen, and Viki Stirling. 2009. Using Measurements from Usability Testing, Search Log Analysis and Web Traffic Analysis to Inform Development of a Complex Web Site Used for Complex Tasks. *Human-Computer Interaction* 10:729–738.

Johnson, Robert R., Michael Salvo, and Meredith W. Zoetewey. 2007. User-Centered Technology in Participatory Culture: Two Decades "Beyond a Narrow Conception of Usability Testing." *IEEE Transactions on Professional Communication* 30:320–332.

Kirakowski, Jurek. 1996. The Software Usability Measurement Inventory, Background and Usage. In *Usability Evaluation in Industry*, ed. Patrick W. Jordan, Bruce Thomas, Bernard A. Weerdmeester, and Ian L. McClelland, 169–178. London: Taylor and Francis.

———. 2005. Summative Usability Testing: Measurement and Sample Size. In *Cost-Justifying Usability: An Update for the Information Age*, ed. Randolph G. Bias and Deborah Mayhew, 519–554. San Francisco: Morgan Kaufmann.

Kirakowski, Jurek, and Bozena Cierlik. 1998. Measuring the Usability of Web Sites. In *Human Factors and Ergonomics Society 42nd Annual Meeting*. Santa Monica, CA: HFES. http://www.ucc.ie/hfrg/questionnaires/wammi/research.html. Accessed February 12, 2010.

Koerber, Amy, and Lonie McMichael. 2008. Qualitative Sampling Methods: A Primer for Technical Communicators. *Journal of Business and Technical Communication* 22:454–473.

Krippendorff. Klaus. 2004. *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks, CA: Sage.

Krug, Steve. 2005. *Don't Make Me Think*. Berkeley, CA: New Riders Press.

Kushniruk, Andre, and Vimla Patel. 2004. Cognitive and Usability Engineering Methods for the Evaluation of Clinical Information Systems. *Journal of Biomedical Informatics* 37:56–76.

Lewis, James R. 2006. Usability Testing. In *Handbook of Human Factors and Ergonomics*, 3rd edition, ed. Gavriel Salvendy, 1275–1316. New York: John Wiley.

Makri, Stephann, Ann Blandford, and Anna Louise Cox. 2008. Using Information Behaviors to Evaluate the Functionality and Usability of Electronic Resources: From Ellis's Model to Evaluation. *Journal of the American Society for Information Science and Technology* 59:2244–2267.

Medlock, Michael, Dennis Wixon, Mick McGee, and Dan Welsh. 2005. The Rapid Iterative Test and Evaluation Method. In *Cost-Justifying Usability: An Update for the Information Age*, ed. Randolph G. Bias and Deborah Mayhew, 489–517. San Francisco: Morgan Kaufmann.

Miles, Matthew, and A. Michael Huberman. 1994. *Qualitative Data Analysis*. Thousand Oaks, CA: Sage.

Mirel, Barbara. 2004. *Interaction Design for Complex Problem Solving: Developing Useful and Usable Software*. San Francisco: Elsevier.

Mirel, Barbara, and Zach Wright. 2009. Heuristic Evaluations of Bioinformatics Tools: A Development Case. In *Proceedings of the 13th International Conference on Human Computer Interaction (HCII-09) Lecture Notes in Computer Science 5610*, 329–338. Berlin: Springer.

Molich, Rolf, Robin Jeffries, and Joseph Dumas. 2007. Making Usability Recommendations Useful and Usable. *Journal of Usability Studies* 2:162–179.

Nielsen, Jakob. 1993. *Usability Engineering*. Boston: Academic Press.

———. 1994. Heuristic Evaluation. In *Usability Inspection Methods*, ed. Jakob Nielsen and Robert L. Mack, 25–62. New York: John Wiley.

Redish, Janice, Rudolphe G. Bias, Robert Bailey, Rolf Molich, Joseph Dumas, and Jared Spool. 2002. Usability in Practice: Formative Usability Evaluations—Evolution and Revolution. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, 885–890. New York: ACM.

Rosenbaum, Stephanie. 2008. The Future of Usability Evaluation: Increasing Impact on Value. In *Maturing Usability*, ed. Effie Law, Ebba Hvannberg, and Gilbert Cockton, 344–378. London: Springer.

Rubin, Jeffrey, and Dana Chisnell. 2008. *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. 2nd edition. New York: Wiley.

Saraiya, Purvi, Chris North, and Karen Duca. 2005. An Insight-Based Methodology for Evaluating Bioinformatics Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 11:443–456.

Sauro, Jeff, and Erica Kirklund. 2005. A Method to Standardize Usability Metrics into a Single Score. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 401–409. New York: ACM.

Sauro, Jeff, and James R. Lewis. 2009. Correlations among Prototypical Usability Metrics: Evidence for the Construct of Usability. In *Proceedings of the 27th International Conference on Human Factors in Computing*, 1609–1618. New York: ACM.

Scotch, Matthew, Bambang Parmanto, and Valerie Monaco. 2007. Usability Evaluation of the Spatial OLAP Visualization and Analysis Tool (SOVAT). *Journal of Usability Studies* 2:76–95.

Snyder, Carolyn. 2003. *Paper Prototyping: The Fast and Easy Way to Design and Refine User Interfaces*. San Francisco: Morgan Kaufmann.

Spyridakis, Jan, Carolyn Wei, Jennifer Barrick, Elisabeth Cuddihy, and Brandon Maust. 2005. Internet-Based Research: Providing a Foundation for Web-Design Guidelines. *IEEE Transactions on Professional Communication* 48:242–260.

Sutcliffe, Alistair, Michele Ryan, Ann Doubleday, and Mark Springett. 2000. Model Mismatch Analysis: Towards a Deeper Explanation of Users' Usability. *Behaviour and Information Technology* 19:43–55.

Sullivan, Patricia. 1989. Beyond a Narrow Conception of Usability Testing. *IEEE Transactions on Professional Communication* 32:256–264.

Theofanos, Mary, and Whitney Quesenbery. 2005. Towards the Design of Effective Formative Test Reports. *Journal of Usability Studies* 1:27–45.

Ummelen, Nicole. 1997. *Procedural and Declarative Information in Software Manuals: Effects of Information Use, Task Performance, and Knowledge*. Amsterdam: Rodopi.

Wilson, Chauncy, and Kara Coyne. 2001. Tracking Usability Issues: To Bug or Not to Bug? *Interactions* 8:15–19.