Introduction: New Perspectives on Document Evaluation Methods

—MENNO DE JONG
AND JUDITH RAMEY,
MEMBER, IEEE,
Guest Editors

Abstract— *Document evaluation and usability testing are gradually developing into an interesting area of multidisciplinary research. Five recent developments are discussed: the broadening notion of document quality, the increasing variety of evaluation approaches, the need to integrate experiences and research from various communication disciplines, the methodological research into evaluation approaches, and the use of evaluation results to formulate design guidelines.*

Eleven years ago, a special issue of IEEE TRANSACTIONS ON PROFESSIONAL COMMUNICATION drew the attention of many technical communicators and researchers to the importance of usability evaluation [1]. The papers in that collection reflected on the nature, possibilities, and practices of usability evaluation approaches. One of the articles was an influential overview of methods by Schriver, who distinguished text-focused, expert-judgment-focused, and reader-focused evaluation techniques, and expressed a clear preference for the last category [2]. For us, the articles in that special issue have been a source of inspiration for a great deal of our research efforts in the 1990s. We have seen other scholars and practitioners exploring the pros and cons of evaluation methods as well. Several special issues of other scientific journals have been dedicated to the topic of usability evaluation in the past decade [3]–[7].

In this special issue, we want to pick up the thread and explore some of the recent developments in the field. As the title of this special issue suggests, we have looked for new perspectives on document evaluation and usability testing. The resulting set of articles addresses research questions that have not been investigated before, at least not in a systematic way. In the remainder of this introduction, we outline what kinds of developments we see in the field of document evaluation and how the articles contained in this special issue reflect these developments.

DEVELOPMENTS IN THE FIELD

In the professional and scientific literature on technical communication, we see several developments regarding document evaluation and usability testing that are indicative of a dynamic and evolving field.

First of all, the overall concept of document quality is expanding from the traditional task-oriented concept of usability to a much greater variety of aspects, comprising, for instance, utility and likability. This can be seen particularly, but not only, in the context of the World Wide Web. Spool et al. [8], for example, found out that the content of websites contributed more to the users' overall appreciation than the aspects of usability that they focused on. Additionally, several studies into Internet use suggest that the traditional Human–Computer Interaction (HCI) assumptions about Web visitors (e.g., the image of people expressly searching for specific

information on a website) may be too narrow [9], [10]. Web designers and evaluators may also have to focus on quality aspects for visitors with more or less exploratory surfing behavior. In addition to the broadening concept of document quality, there is also a growing attention to the multicultural and multilingual aspects of documents [11]. New evaluation methods and designs will have to address these new requirements.

Partly corresponding to the expanding concept of document quality, many new evaluation techniques have been and are being developed. Developments can be seen in various directions [12], [13]. In addition to task-oriented usability test methods, nonspecific evaluation methods, such as the plus–minus method, have been developed, described, and investigated [14], [15]. There has also been a re-evaluation of text-focused and expert-judgment-focused evaluation methods. In 1989, these methods were heavily criticized by Schriver in favor of reader-focused alternatives [2]. In the 1990s, however, several researchers have taken up the challenge of developing analytical evaluation tools that would help experts to give valid feedback on interfaces and documents, preferably from the users' point of view [16]–[18]. Especially heuristic evaluation, cognitive walkthroughs, and functional analyses seem to be promising techniques, complementing rather than competing with reader-focused methods. An entirely different development regarding methods is the growing attention for more or less ethnographic approaches of evaluation research. Instead of inviting participants to carry out a set of predetermined tasks in a laboratory setting, approaches have been developed in which (potential) users are observed and interviewed in their natural work setting [19], [20]. Finally, the possibilities of collecting data

by computer about the usability of documents and interfaces are being explored. Several kinds of rapid, remote, and automated tools have been and are being developed. Remote usability evaluation is possible using the Internet: potential users can participate in a usability test without going to a usability lab at a preset time. Software can be used to record the participants' comments and log their process of using the document or interface [21], [22]. Another interesting option is called "kiosk-based user testing" [23]. This is an approach that can be used to engage participants in events like conferences in a usability test. On a stand-alone computer, an entire usability evaluation procedure is programmed, and volunteers can participate without the help of a facilitator. All in all, the "methodological toolkit" of technical communicators and usability professionals can be expected to contain a much greater variety of techniques than ten years ago.

A third development is that the borders between different communication disciplines studying evaluation methods are gradually disappearing. Originally, very few cross-references could be found between technical communication, interface design, questionnaire design, instructional design, and public information, although the problem of evaluating artifacts has much in common in all disciplines. In our view, it may be very helpful for technical communicators to further explore the developments in document evaluation in other disciplines. In addition to the popular handbooks on usability testing (see [24]–[26]), practitioners could benefit as well from relevant publications on, for instance, interface design [27], [28] and instructional design [29]–[31]. The same applies to the methodological research into the merits and restrictions of various evaluation approaches. In general terms, the research problem is

similar in each discipline: what are the relevant questions to be asked about the value of evaluation techniques, and what are the best research designs for answering these questions? Although some caution with respect to the transferability of results is of course advisable, the results found in other disciplines may be quite informative.

A fourth development is that the methodology of document evaluation and usability testing has gained a place in the academic research agenda. In Schriver's overview on document evaluation in 1989, the majority of the empirically based references were borrowed from research areas other than document evaluation [2]. Now, more specific research is available, varying from case studies about the practice of evaluating documents to the experimental comparison of the yield of different methods. Given the enormous variety of evaluation options, we hope that the methods of document evaluation and usability testing will soon be even higher on the research agenda.

Finally, several attempts can be seen to systematically use evaluation results to gain new insight in communication processes. Several researchers have explored whether the results of usability evaluation can also be used for more general purposes than just improving the artifact evaluated. Schulz et al. [32] carried out a post hoc analysis of usability test results to explore the relationship between user-generated and designer-generated metaphors. They discovered that users often stick to the metaphors they bring to the task and get confused if there is a mismatch between user-generated and designer-generated metaphors. Bachiochi et al. [33] describe how usability testing was used to decide on the optimal navigational aids for websites. Other publications present various design guidelines

based on the results of a series of usability tests [8], [34], [35]. Although the step from test results to guidelines is not explicitly accounted for in these publications, the basic idea of trying to translate usability test results in design guidelines seems to be very useful.

ABOUT THE ARTICLES

In this special issue, we have collected a variety of articles, reflecting many of the developments described above. The variety of the articles concerns not only the evaluation methods addressed and the artifacts the methods were used on, but also the research questions about the evaluation methods and the types of research carried out to answer these questions. Together, the six articles highlight various aspects of document evaluation and usability testing, both for practitioners who want to include serious evaluation activities in their design processes and for researchers who are interested in joining us in future research efforts in this area.

The first contribution is a state-of-the-art overview. De Jong and Schellens present an outline of the available research on the topic of document evaluation (the other contributions in this special issue not included). The overview is restricted to empirical studies actually putting the merits and restrictions of evaluation approaches to the test. They expressly try to integrate research stemming from various disciplines, such as technical communication, instructional design, questionnaire design, public information, as well as interface and Web design. The overview can serve as an inventory of what is empirically known about document evaluation and as the beginning of a research agenda.

In the area of technical communication, think-aloud user protocols are one of the most widely used current approaches. Two of the articles in this special

issue focus on the design of this type of usability testing. Their merit is that they discuss aspects of usability testing that often seem to be taken for granted in practice, in handbooks, and in the research literature.

Boren and Ramey focus on the role of the facilitator during usability test sessions. They discuss the tenability of the classical theoretical assumptions about verbal reports by Ericsson and Simon [36] as a basis for think-aloud usability testing. For a long time, usability professionals seem to have relied on this classic work to justify their test practices. Boren and Ramey observed in a field study that Ericsson and Simon's guidelines for think-aloud research are often violated in practice. Instead of immediately blaming usability professionals for irresponsible behavior, they explore whether the strict guidelines for think-aloud research are indeed applicable to usability testing and propose an alternative "speech communication" approach.

Van Waes discusses another design variation of think-aloud user protocols. He focuses on the tasks given to the participants in a usability test. Van Waes distinguishes between three types of tasks—simple searching, application, and prediction—and shows in a qualitative analysis of think-aloud data that each task type may lead to the detection of different types of user problems.

The fourth contribution to this special issue compares the results of think-aloud user protocols to those of an alternative evaluation approach. Ummelen and Neutelings investigate the "reactivity" of alternative ways to track the selection behavior of documents users (i.e., the extent to which both methods affect the ways participants behave and perform in a test situation). In this study, think-aloud user protocols are experimentally compared to

the "click & read method" that Ummelen [37] originally developed to gain insight in the extent to which participants use procedural and declarative information in manuals.

The remaining two articles in this special issue are case studies focusing on two entirely different aspects of document evaluation. Both case studies address issues that are current in technical communication, though neither has received much attention in the research.

Evans describes the development of a set of research-based guidelines for websites. Especially for the World Wide Web, many guidelines and heuristics are available, but the empirical foundation of the advice contained in them is often unclear [38]. Evans tries to formulate guidelines based explicitly on the results of empirical research. Her case study demonstrates that the process of developing such research-based guidelines is problematic and far from straightforward. Based on her own experiences, Evans explains what a more considered guideline development process might look like.

Finally, Lentz and Hulst present a case study about the evaluation of multilingual documents. They describe an extensive document evaluation project in which they combined expert-judgment-focused and reader-focused methods to test various language versions of a public information brochure. Many of the difficulties they faced concerned the overall problem of "cognitive distribution" (i.e., the required knowledge may be available in an organization, but it is divided among different persons who have trouble communicating their experiences to each other and who cannot control each other's activities).

Together, the articles in this special issue provide a diverse overview

of the work in progress in this fascinating area of research. Much progress has been made in the past decade, but many questions still exist. The growing body of research does not necessarily make things easier for practitioners who want to carry out a well-considered usability test. On the contrary, research may urge us to reconsider established evaluation practices that were deemed unproblematic or to take a more nuanced position than before. Ten years of scholarly research have, for instance, added many *ifs* and *buts* to Schriver's clear-cut preference for reader-focused methods [2]. But eventually, research attention to and reflection on evaluation methods will prove to be a precondition for the effective use of evaluation research in document design processes. We hope that the articles in this special issue will incite other researchers and practitioners to pursue new research efforts in this area.

REFERENCES

[1] J. Ramey, Ed., "Usability evaluation and its relationship to traditional approaches in research and design," *IEEE Trans. Prof. Commun.*, vol. 32, pp. 207–316, 1989.

[2] K. A. Schriver, "Evaluating text quality: The continuum from text-focused to reader-focused methods," *IEEE Trans. Prof. Commun.*, vol. 32, pp. 238–255, 1989.

[3] J. Nielsen, Ed., "Usability laboratories," *Behav. Inform. Technol.*, vol. 13, pp. 3–197, 1994.

[4] J. Renkema and P. J. Schellens, Eds., "Tekstevaluatie (Text evaluation)," *Taalbeheersing*, vol. 18, pp. 305–392, 1996.

[5] P. J. Schellens and T. van der Geest, Eds., "Formative evaluation of professional documents," *J. Bus. Tech. Commun.*, vol. 11, pp. 399–501, 1997.

[6] D. L. Scapin and T. Berns, Eds., "Usability evaluation methods," *Behav. Inform. Technol.*, vol. 16, pp. 185–293, 1997.

[7] G. M. Olson and T. P. Moran, Eds., "Experimental comparisons of usability evaluation methods," *Human-Comput. Interact.*, vol. 13, pp. 199–335, 1998.

[8] J. M. Spool, T. Scanlon, W. Schroeder, C. Snyder, and T. DeAngelo, *Web Site Usability: A Designer's Guide*. North Andover, MA: User Interface Eng., 1997.

[9] S. A. Barab, B. E. Bowdish, and K. A. Lawless, "Hypermedia navigation: Profiles of hypermedia users," *Educ. Technol. Res. Devel.*, vol. 45, no. 3, pp. 23–41, 1997.

[10] S. Muylle, R. Moenaert, and M. Despontin, "A grounded theory of World Wide Web Search behavior," *J. Market. Commun.*, vol. 5, pp. 143–155, 1999.

[11] N. L. Hoft, *International Technical Communication: How to Export Information About High Technology*. New York: Wiley, 1995.

[12] P. W. Jordan, B. Thomas, B. A. Weerdmeester, and I. L. McClelland, Eds., *Usability Evaluation in Industry*. London, U.K.: Taylor & Francis, 1996.

[13] K. Halskov Madsen, Ed., "The diversity of usability practices," *Commun. ACM*, vol. 42, no. 5, pp. 61–97, 1999.

[14] M. de Jong and P. J. Schellens, "Reader-focused text evaluation: An overview of goals and methods," *J. Bus. Tech. Commun.*, vol. 11, pp. 402–432, 1997.

[15] M. de Jong, *Reader Feedback in Text Design: Validity of the Plus–Minus Method for the Pretesting of Public Information Brochures*. Atlanta, GA: Rodopi, 1998.

[16] J. Nielsen and R. L. Mack, *Usability Inspection Methods*. New York: Wiley, 1994.

[17] L. Lentz and H. Pander Maat, "Evaluating text quality: Reader-focused or text-focused?," in *Studies of Functional Text Quality*, H. Pander Maat and M. Steehouder, Eds. Atlanta, GA: Rodopi, 1992, pp. 101–114.

[18] T. van der Geest and J. Spyridakis, Eds., "Web communication heuristics," *Tech. Commun.*, vol. 47, to be published.

[19] D. Wixon and J. Ramey, Eds., *Field Methods—Casebook for Software Design*. New York: Wiley, 1996.

[20] P. Dorazio and J. Stovall, "Research in context: Ethnographic usability," *J. Tech. Writing and Commun.*, vol. 27, pp. 57–67, 1997.

[21] D. R. Millen, "Remote usability evaluation: User participation in the design of a Web-based email service," *SIGGroup Bull.*, vol. 20, no. 1, pp. 40–45, 1999.

[22] J. Scholtz, S. Laskowski, and L. Downey, "Developing usability tools and techniques for designing and testing Web sites," in *Proc. 4th Conf. on Human Factors & the Web*, Baskerville, NJ, June 5, 1998, [Online]. Available: http://www.research.att.com/conf/hfweb/proceedings/scholtz.

[23] J. Scholtz and S. Laskowski, "Kiosk-based user testing of online books," in *Proc. 16th Annu. Int. Conf. Computer Documentation*. Quebec, Canada, Sept. 23–26, 1998, pp. 80–86.

[24] J. S. Dumas and J. C. Redish, *A Practical Guide to Usability Testing*. Norwood, NJ: Ablex, 1993.

[25] J. Rubin, *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. New York: Wiley, 1994.

[26] C. Velotta, Ed., *Practical Approaches to Usability Testing for Technical Documentation*. Arlington, VA: STC, 1995.

[27] J. Nielsen, *Usability Engineering*. Boston, MA: Academic, 1993.

[28] G. Lindgaard, *Usability Testing and System Evaluation: A Guide for Designing Useful Computer Systems*. London, U.K.: Chapman & Hall, 1994.

[29] B. N. Flagg, *Formative Evaluation for Educational Technologies*. Hillsdale, NJ: Lawrence Erlbaum, 1990.

[30] M. Tessmer, *Planning and Conducting Formative Evaluations: Improving the Quality of Education and Training*. London, U.K.: Kogan Page, 1993.

[31] J. George and J. Cowan, *A Handbook of Techniques for Formative Evaluation: Mapping the Student's Learning Experience*. London, U.K.: Kogan Page, 1999.

[32] E. Schulz, J. Ramey, M. van Alphen, and W. Rasnake, "Discovering user-generated metaphors through usability testing," *IEEE Trans. Prof. Commun.*, vol. 40, pp. 255–264, Dec. 1997.

[33] D. Bachiochoi, M. Berstene, E. Chouinard, N. Conlan, M. Danchak, T. Furey, C. Neligon, and D. Way, "Usability studies and designing navigational aids for the World Wide Web," *Comput. Networks ISDN Syst.*, vol. 29, pp. 1489–1496, 1997.

[34] J. Nielsen, "Heuristic evaluation," in *Usability Inspection Methods*, J. Nielsen and R. L. Mack, Eds. New York: Wiley, 1994, pp. 25–62.

[35] J. A. Borges, I. Morales, and N. J. Rodríguez, "Page design guidelines developed through usability testing," in *Human Factors and Web Development*, C. Forsythe, E. Grose, and J. Ratner, Eds. Mahwah, NJ: Lawrence Erlbaum, 1998.

[36] K. A. Ericsson and H. A. Simon, *Protocol Analysis: Verbal Reports as Data (Revised Edition)*. Cambridge, MA: MIT Press, 1993.

[37] M. M. N. Ummelen, *Procedural and Declarative Information in Software Manuals: Effects on Information Use, Task Performance and Knowledge*. Atlanta, GA: Rodopi, 1997.

[38] M. de Jong and T. van der Geest, "Characterizing Web heuristics," *Tech. Commun.*, to be published.