



A test-first view of usability

Judy Kay

Professor of Computer Science, CHAI – Computer Human Adapted Interaction Research Group, School of Information Technologies, University of Sydney, Australia

ARTICLE INFO

Article history:

Available online 22 April 2009

Keywords:

Ergonomics
Pervasive
Test-first
Usability evaluation

ABSTRACT

The foundations of usability evaluation were being established in the early 1990s. In this context, “Usability – Context, Definition, Design and Evaluation” built upon Brian Shackel’s earlier influential work – work that helped define the notion of usability. In this paper, he established key dimensions of usability as well as approaches to integrating the testing of these dimensions, within the whole process of setting requirements. Essentially he argued for usability design as part of the system design process.

This commentary describes the context of Professor Shackel’s paper and reviews the influential ideas that appear in much subsequent work.

Crown Copyright © 2009 Published by Elsevier B.V. All rights reserved.

1. Introduction

“Usability – Context, Definition, Design and Evaluation” is a quite short paper. It is almost tutorial-style, defining a quite broad view of usability and then operationalising it to provide foundations for systematic evaluation processes. The paper illustrates ideas in terms of a readily understood, though dated, example.

In today’s terminology, we can see that it argues strongly for a test-first view of usability, with test design integrated into other aspects of the design process, and usability evaluation at the fore from the very beginning. This contrasts with the Waterfall sequential model (TechRepublic, 2009), where all thought of testing was deferred to late in the software process.

The paper points to the benefits of tight usability requirements, for example setting specific, concrete times for task completion and limits on the numbers of errors that are acceptable. Professor Shackel makes the case that these are critical in informing the design process including evaluation. He emphasises the importance of affective aspects and the need for rigor in formulating the evaluation of likeability. In summarising the state of the art in evaluation techniques, he points out the need for progress in defining techniques that could attain effective applicability beyond large-scale, high cost systems.

Notably, the character of the usability requirements suggested in the paper necessitates formal empirical evaluation. They are not amenable to discount non-user approaches, such as expert review approaches (mentioned in the paper). Moreover, since they employ quantitative measures, they would seem to demand large-scale, inherently expensive user studies.

2. HCI in 1991

The computing landscape has changed markedly since 1991, especially for the average citizen. The Internet was just becoming available for private users. It was well before the web, which only began to have a major impact for many people in the late 1990s. This time was before the availability of laptops: computers were expensive workplace mainframes, mini-computers and a growing market of personal computers. All were expensive in real terms.

Computer screens were black-and-white or green. It was before the now ubiquitous mobile phones that are pointing the way to a future where we will carry powerful computational devices – devices that both hold important personal information and enable us to use online services and access information flexibly (as well as supporting easy communication). Computers were becoming increasingly embedded in the fabric of the workplace.

Professor Shackel already had played a key role in the definition of usability (Shackel, 1984, 1986). Even in 1991, there were tensions between the empirical evaluations he advocated for assessing usability, versus lower cost approaches (Thimbleby, 1994). The early 1990s saw a burst of research activity and significant progress in usability assessment, and a shift to an engineering view of usability (Nielsen and Bellcore, 1992). The need for iterative test and development cycles was also argued by Nielsen and Bellcore (1992).

This period saw the birth of the now standard approaches to usability evaluation. Notably, these include the discount methods, such as systematic usability inspection methods (Nielsen, 1994a,b), including heuristic evaluation (Nielsen and Molich, 1990). These were based on established criteria that could reduce somewhat the demands on expertise for effective evaluations. The cognitive walkthrough (Lewis et al., 1990) emerged for low

E-mail address: judy@it.usyd.edu.au

cost evaluations of systems for novice users. Think-aloud approaches had been proposed far earlier (Lewis, 1982).

This period saw active study. For example, Wright and Monk (1991) showed the value of such approaches in overcoming programmer blindness to usability problems, and how programmers could help in the refinement of the interface. In the early 1990s, usability evaluation moved beyond research papers to practitioner books such as the highly influential Nielsen (1994a) book on usability engineering and the Rubin and Hudson (1994) handbook of usability testing. These were followed by a series of such textbooks for students and interface developers such as that by Dumas and Redish (1999). By the mid-1990s, there was a substantial body of work, as surveyed by John (1996).

There was on-going debate on the relative value of discount usability techniques compared with empirical methods, with considerable experimental evidence pointing to the greater power of the latter. (For example, Virzi (1992), Jeffries and Desurvire, 1992), and Karat et al. (1992) all showing the benefits of empirical evaluation approaches suggested in Professor Shackel's paper.)

3. Ideas from this paper in subsequent and emerging HCI

When Professor Shackel wrote this paper, he contributed memes to the software and HCI community that promoted designing for usability and integrating HCI evaluation into the design process. He clearly was influenced by the existing broad understandings of these issues, and contributed to the evolving understandings of them. This section outlines some of the key themes of his paper that appeared in subsequent work.

3.1. Test-first and iterative design

One striking feature of Professor Shackel's paper is his argument for a test-first approach, integrated into a series of design iterations. He points to the merits of this approach in informing the design process – as well as providing a framework for the assessment of the various aspects of usability. This is offered in the spirit of one of the tenets of extreme programming (Beck, 1999) and other Agile approaches (Cockburn, 2002) to software development.

There has been much work to assess the power of such approaches to reduce software defects (Williams et al., 2003). Test-first approaches also were shown to improve the effectiveness of students (Erdogmus, 2005), in terms of productivity and quality. But the picture is not simple. For example, such approaches may not increase the speed of implementation or reliability of the programs, but may improve program understanding. Such process issues are at the core of Professor Shackel's paper.

3.2. On-going research into usability test methods

Despite all the progress of the early 1990s, there is still important work to be done. Some aimed to clarify the relative power of different techniques, providing insights into the reasons for their relative strengths. Other work sought to make techniques more reliable and robust, even in the hands of the many programmers who build interfaces without support from HCI experts. For example, a decade after Professor Shackel's paper, a comparison of the widely used cognitive walkthrough, heuristic evaluation, and think-aloud demonstrated a strong evaluator effect. Or, taking the case of the think-aloud, there are echoes of Professor Shackel's view of the critical role of explicit preliminary “descriptions of design, test procedure and framework for analysis” (Hertzum and Jacobsen, 2003).

Even in the most recent HCI conferences, there has been a strong focus on usability evaluation. For example, the potentially

contentious paper by Greenberg and Buxton (2008) would have pleased Professor Shackel in its call for mindful (rather than rule-based) usability evaluations. His paper makes it clear that it is important to account for the context of use. This notion has figured in standards and in considerable work, such as reviewed in Maguire (2001).

3.3. The web and consumer interfaces and personalisation

The emergence of the web has radically changed the ways that the broad population can make use of computers. There has been a corresponding growth in web-oriented usability guidelines, tools, standards and techniques.

One of the important differences between web-based interfaces and the workplace systems which dominated the early 1990s is the increasing importance of the *likeability* of the interface, as well as its utility. These are reflected in work by Gaines et al. (1996) for the web, Zviran et al. (2006) for commercial web sites, Zaman (2008) for children's games, and Han et al. (2001) in relation to consumer electronic products. The web also opens many new possibilities for usability testing – for example, the Mechanical Turk's Human Intelligence Tasks.¹

3.4. Dynamic or personalised usability

A key part of my user modeling and personalisation research has attempted to create an explicit model of the user built from a range of evidence. Core design goals are that users should be able to scrutinise and have control – control of their own models and the processes that form them (Kay, 1995). This is important, especially in the case of pervasive computing systems (Assad et al., 2007).

One of the earliest uses for such models was inspired by Professor Shackel's operational view of usability. Essentially, the user model can serve as a source of hard evidence about the ways that each individual actually uses an application, as well as the ways that this usage evolved over time (Cook et al., 1995). This introduces the notion of *personal usability*, acknowledging that each person's needs, context and likes will affect usability for them.

4. Conclusions

“Usability – Context, Definition, Design and Evaluation” called for careful design – the kind of design that integrates thoughtful identification of criteria for success in the various dimensions of usability, and rigor in defining the mechanisms for measuring it.

Professor Shackel clearly recognised the immaturity of mechanisms for these evaluations. He would have been pleased to see how seriously the HCI community has taken the call for ways to improve our understanding of usability and its measurement. He also would have been pleased to see that usability and HCI have been recognised as an essential part of the curriculum of computing degrees.²

Professor Shackel's 1991 paper hints at the tension between discount usability and more costly, larger-scale testing required to assess usability requirements in the examples of his paper. But, his strongest call was for thoughtful design that takes account of the broad dimensions of usability, integrated with definition of measurable usability requirements. This approach would be an antidote to sloppy, vague, and wishful assessment of usability. It is important today, but we still have a way to go in learning how to do it well and to see it become the norm.

¹ <https://www.mturk.com/mturk/welcome>.

² <http://www.acm.org/education/curricula-recommendations>.

About the author

Judy Kay is Professor of Computer Science at the School of Information Technologies at the University of Sydney. She leads the Computer Human Adapted Interaction Research Group (CHAI), conducting research in advanced technologies for human computer interaction, supporting personalisation, pervasive and mobile interaction. Her vision is to support a highly personalised way for people to interact with the computers that they carry, those embedded within the environment as well as desktop computers. Critical to this vision is that the user should always feel a sense of control over the personalisation. She creates new tools for interaction in support of lifelong learning, collaboration and personalised smart services.

Professor Kay has extensive publications in the areas of personalisation, user modelling and advanced teaching systems. She has given invited keynote addresses at conferences such as UM'94 (User Modeling), IJCAI'95 (International Joint Conference on Artificial Intelligence), ICCE'97 (International Conference on Computers in Education), ITS'2000 (Intelligent Tutoring Systems), AH2006 (Adaptive Hypermedia and Adaptive Web-Based Systems), ITS2008 (Intelligent Tutoring Systems).

Professor Kay holds several editorial positions:

- Associate Editor, Artificial Intelligence in Education;
- Associate Editor, IEEE TLT Transaction on Learning Technologies;
- Editorial Board, UMUAI (User modeling and User-Adapted Interaction: the Journal of Personalization Research)

She is President-elect for the AIED Society (Artificial Intelligence in Education) and is on the steering committee for ITS (Intelligent Tutoring Systems) and UM Inc (User Modeling).

Professor Kay may be contacted at judy@it.usyd.edu.au.

References

- Assad, M., Carmichael, D.J., Kay, J., Kummerfeld, B., 2007. PersonisAD: distributed, active, scrutable model framework for context-aware services. In: Proceedings of Pervasive 07 5th International Conference on Pervasive Computing. Lecture Notes in Computer Science, vol. 4480. Springer, pp. 55–72.
- Beck, K., 1999. Extreme Programming Explained: Embrace Change. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA.
- Cockburn, A., 2002. Agile Software Development. Addison-Wesley Boston.
- Cook, R., Kay, J., Ryan, G., Thomas, R., 1995. A toolkit for appraising the long-term usability of a text editor. *Software Quality Journal* 4 (2), 131–154.
- Dumas, J., Redish, J., 1999. A Practical Guide to Usability Testing. Intellect Books.
- Erdogmus, H., 2005. On the effectiveness of the test-first approach to programming. *IEEE Transactions on Software Engineering* 31 (3), 226–237 (Member-Maurizio Morisio and Member-Marco Torchiano).
- Gaines, B., Shaw, L., Chen, L., 1996. Utility, Usability and Likeability: Dimensions of the Net and Web. Proc of WebNet96, San Francisco, CA, AACE. <<http://ksi.cpsc.ucalgary.ca/articles/WN96/WN96HF/WN96HF.html>> (03.08.04).
- Greenberg, S., Buxton, B., 2008. Usability evaluation considered harmful (some of the time). In: CHI '08: Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 111–120.
- Han, S., Yun, M., Kwahk, J., Hong, S., 2001. Usability of consumer electronic products. *International Journal of Industrial Ergonomics* 28 (3–4), 143–151.
- Hertzum, M., Jacobsen, N., 2003. The evaluator effect: a chilling fact about usability evaluation methods. *International Journal of Human-Computer Interaction* 15 (1), 183–204.
- Jeffries, R., Desurvire, H., 1992. Usability testing vs. heuristic evaluation: was there a contest? *SIGCHI Bulletin* 24 (4), 39–41.
- John, B.E., 1996. Evaluating usability evaluation techniques. *ACM Computing Surveys*, 139.
- Karat, C.-M., Campbell, R., Fiegel, T., 1992. Comparison of empirical testing and walkthrough methods in user interface evaluation. In: CHI '92: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 397–404.
- Kay, J., 1995. The um toolkit for cooperative user modelling. *User Modeling and User-Adapted Interaction* 4, 149–196.
- Lewis, C., 1982. Using the "Thinking-Aloud" METHOD in Cognitive Interface Design. IBM TJ Watson Research Center.
- Lewis, C., Polson, P.G., Wharton, C., Rieman, J., 1990. Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. In: CHI '90: Proceedings of the SIGCHI Conference on HUMAN Factors in Computing Systems. ACM, New York, NY, USA, pp. 235–242.
- Maguire, M., 2001. Context of use within usability activities. *International Journal of Human Computer Studies* 55 (4), 453–484.
- Nielsen, J., 1994a. Usability Engineering. Morgan Kaufmann.
- Nielsen, J., 1994b. Usability inspection methods. In: Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 413–414.
- Nielsen, J., Bellcore, M., 1992. The usability engineering life cycle. *Computer* 25 (3), 12–22.
- Nielsen, J., Molich, R., 1990. Heuristic evaluation of user interfaces. In: CHI '90: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 249–256.
- Rubin, J., Hudson, T., 1994. Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests. John Wiley & Sons, Inc., Berlin.
- Shackel, B., 1984. The concept of usability. *Visual Display Terminals: Usability Issues and Health Concerns*, 45–87.
- Shackel, B., 1986. Ergonomics in design for usability. In: Conference of British Computer Society Human Computer Interaction Specialist Group. Cambridge University Press, pp. 44–64.
- TechRepublic, 2009. Understanding the pros and cons of the waterfall model of software development. <http://articles.techrepublic.com/5100-10878_11-6118423.html> (accessed 11.03.09).
- Thimbleby, H., 1994. Formulating usability. *SIGCHI Bulletin* 26 (2), 59–64.
- Virzi, R., 1992. Refining the test phase of usability evaluation: how many subjects is enough? *Human Factors: The Journal of the Human Factors and Ergonomics Society* 34 (4), 457–468.
- Williams, L., Maximilien, E. M., Vouk, M., 2003. Test-driven development as a defect-reduction practice. In: International Symposium on Software Reliability Engineering, pp. 34–45.
- Wright, P.C., Monk, A.F., 1991. The use of think-aloud evaluation methods in design. *SIGCHI Bulletin* 23 (1), 55–57.
- Zaman, B., 2008. Introducing contextual laddering to evaluate the likeability of games with children. *Cognition, Technology & Work* 10 (2), 107–117.
- Zviran, M., Glezer, C., Avni, I., 2006. User satisfaction from commercial web sites: the effect of design and use. *Information & Management* 43 (2), 157–178.