

How to compare usability testing methods with children participants

Panos Markopoulos & Mathilde Bekker

Technology Management, Technical University of Eindhoven
Den Dolech 2, 5600 MB Eindhoven, The Netherlands
P.Markopoulos@tue.nl, M.M.Bekker@tue.nl

Abstract. The paper motivates the need to acquire methodological knowledge for involving children as test users in usability testing. It summarises a methodological framework for comparing usability tests and discusses its implications for future research.

1 Introduction

Usability testing is a cornerstone activity for usability engineering and user centred design. It seems contrived to argue that products intended by children should be tested by children. However, a recent review of current practice in software testing for children by Bruckleitner (1999) observed that most software testing is done by adult experts and seldom are such evaluations related to opinions of children users.

In the field of human computer interaction, a growing interest in children as technology users has lead to efforts to encode methodological knowledge for usability testing with children as test participants. This type of work is very sparse. Hanna, Ridsen and Alexander (1997) provide the most comprehensive treatment of the subject to date. They describe a set of guidelines for usability testing with children in the laboratory that is based upon their own practical experience. This set of guidelines touches on a wide diversity of subjects, such as the decoration of the room, the observation equipment used (e.g., use small microphones), the behaviour of the experimenter (how to encourage them when they fail at a task), etc.

Our research aims to derive a systematic account of how the different and changing capacities and characteristics of children affect usability testing. This research aims to derive methodological knowledge for prescribing how to conduct usability testing with children participants. Several questions need to be addressed to meet this objective: What makes children special as testers of technology? What capabilities/characteristics of them need to be taken into account? This paper presents a framework that provides some structure to such questions and suggests directions for empirical research.

2 A framework for comparing testing methods

This section describes a framework for characterising comparative studies of Usability Testing Methods (UTM) with respect to their appropriateness for children. The methodological framework for comparing UTM's has three dimensions:

1. Performance descriptors/indicators for UTM techniques.

2. Parameters for shaping a UTM set-up.
3. Characteristics of children that potentially influence how a UTM performs.

2.1 Performance Indicators for UTM's

There are several ways to characterise methods for the benefit of practitioners. These are the feasibility of applying the method in a particular content (robustness), the validity of its results, how effective it is in meeting its aim and how efficient use of resources is made. Here, we discuss the meaning of these concepts in the context of usability testing and why some aspects are more salient when we discuss usability testing with children.

The **validity** of usability testing methods pertains to whether the problems it helps uncover are actually usability problems or not. There is currently no common understanding of what constitutes a usability problem and what not. Evidence towards the validity of a problem report can be obtained by identifying the same problem through different usability evaluation techniques.

Reliability pertains to what extent do factors external to the definition of a method influence its outcome. The reliability of a usability testing method concerns whether the individuals acting as test participants, facilitators or evaluators influence the outcome of the method when the same testing procedure is used for the same product? Can other contextual factors (e.g., who is in the room, the setting of the laboratory, etc.) influence the outcome of a method? The reliability of usability methods has been studied extensively particularly with respect to different individuals as test participants (e.g., Nielsen & Landauer 1993).

The **effectiveness** of a usability test can be described in terms of several parameters: the number of problems found, the number of problems missed, the number of false alarms, and in each of these cases the severity of problems concerned. In the absence of a standard definition of what constitutes a usability problem, a broad-brush alternative which many comparative studies adopt is to count all problems reported by test participants.

Efficiency pertains to the amount of the resources used to execute a particular usability in relation to its outcome, e.g., the time required for a test session, the number of real problems found per child participant or evaluator. The studies of Nielsen (1995) and Jacobsen (1999) are good examples of models of the efficiency of a method with respect to the number of test participants and in the latter case of usability evaluators.

The **robustness** of a method concerns the feasibility of applying it to different contexts, such as the laboratory versus the field, work related products versus entertainment related ones, early in product development versus late, etc. Children are not a homogeneous user group. Robustness pertains to different ages or capacities of children that may render inapplicable a method established for adult participants.

2.2 Parameters for describing usability testing method

This section enumerates parameters along which we can vary our set up for usability testing. Ideally, a fairly accurate description of any usability testing method should be obtained by describing it in terms of these parameters. This set of parameters helps systematise the collection of usability testing approaches found in the literature, and to

comprehend the comparisons between them. The list of parameters we use so far is as follows:

- **Purpose of the test.** E.g., to evaluate a preliminary concept, to find ‘bugs’ that must be fixed, to benchmark against competition.
- **The artefact tested.** The artefact may be simple or complex, it may still be in the form of a preliminary prototype or a final product.
- **The interaction tasks.** During a usability-test participants are given a set of tasks to perform. These may be abstract, e.g., type-in a short sentence, or concrete, e.g. type-in “ABCDEF”. Tasks can be described by their complexity, their abstractness, their number and duration, whether they are exploratory or goal-oriented.
- **Participants.** UTM’s are characterised by the number of participants involved, how representative they are of the target user group, etc.
- **Facilitator.** The facilitator runs the usability test. We distinguish this role from the one of the evaluator, which is the one who analyses the data, although these distinct roles can be assumed by the same individual.
- **Environment/context.** Is it in the laboratory, the intended context of use environment? Is it the home or the classroom? Is this place noisy, are there more people around, or is it isolated?
- **Procedure.** Do subjects work alone, in pairs, or in a group? Does the facilitator interact with them, observe them, interview them, prompt them? How? What instructions are given? When? In what form? Are users trained first?
- **Capture of data.** What technique is used to capture empirical data? Is capturing of data interfering with the primary tasks? Are they explicit tasks by the user or is the user observed? Is data collected by interviews, logging, video-recording or note-keeping?

2.3 Characteristics of children that impact the usability test

Characteristics of children that may affect the process of usability testing and its outcome are the following:

- Children have developing capacity to *verbalise (vocally or in writing)*. A related skill is the ability to think aloud. This requires the child to translate their experiences to verbal statements.
- Some children may not be used to speaking up to adults and may be less likely to report usability problems. Different degrees of *extroversion* skills may have a direct impact on the outcome of the usability test.
- Children have developing *capability to concentrate* to a single activity and to pursue tasks. The impact on the usability test is that for different age groups different 'sizes of tasks should be given. Hanna et al. (1999) mention that children can concentrate for about 30 minutes, though in our studies we have found that children who are enjoying themselves have no problem with participating in a longer session.

- Frequent facilitator intervention together with high *motivation* to please adults (evident at younger ages), could also discriminate usability testing procedures (Hanna et al., 1997).
- *Prior experience with computing* should not be ignored in usability studies with children. Wartella (2000) suggests that owners of a computer demonstrate more positive attitudes, more enthusiasm and report more self-confidence and ease when using computers than those who do not.
- *Trustworthiness of self-report*. Children are reported to be very honest (Druin, 1999) but sometimes the reliability of reported data is questionable (Hanna et al. 1999): Children may say they hate the ‘bad’ character, that may in fact be crucial to the success of a product. Alternatively, they may simply name problems to please the evaluator who looks for them or they may conceal problems if they think they would offend the software creators. Depending on their age children are influenced by others, such as other children, parents and teachers in different ways (Acuff, 1997).
- Children develop the *ability for abstract and logical thinking* over time (Berk, 1997). They also become better at doing more complex reasoning, such as cause and effect reasoning. This influences their ability to understand abstract task descriptions and abstract questions for feedback. It also influences their ability to compare or judge products on multiple aspects, varying in level of detail. Furthermore, it influences the number of concepts they can keep in mind at once (Greig and Taylor, 1999) or their ability to score their preferences/opinions along Likert scales.
- Children slowly develop the ability to monitor *progress towards a goal* over childhood and adolescence (Berk, 1997). They have to build up the ability to check outcomes of a task, and learn to redirect unsuccessful efforts. The level of this ability will influence how good they are at conducting goal directed tasks and assessing whether they have done them correctly.
- *Gender differences* develop and change as children become older (Berk, 1997). One example of how gender differences may influence usability tests is from a study with children between 9 and 11 years old, where girls were more verbose than boys, and girls provided more arguments for their opinion (Beusmans, 2002).
- The way in which products are presented can influence the manner in which they are evaluated. According to Acuff (1997), it is important to show 3-dimensional objects rather than flat representations to children.
- *Motor skills* may influence the outcome of a usability evaluation. For example, many children younger than 3 simply cannot use standard input devices like the mouse.

3 Discussion

The framework introduced allows comparative studies of UTM’s to be described along the first two dimensions and potential complications resulting from children-participants are covered by the third. The purpose of this framework is to help generate interesting

hypotheses that can be validated experimentally and to gain a grip on the overall topic rather than focus on point-solutions.

To date, we have carried out two empirical studies within the dimension space of this framework:

Study 1. The first study hypothesized that children with higher verbalisation and extroversion skills will report more usability problems when the usability testing procedure requires active verbalisation skills, e.g., thinking aloud rather than filling in written questionnaires. This hypothesis was not confirmed (cf. Donker and Markopoulos 2002). The results of the experiment suggest that think-aloud helps generate more problem reports than questionnaires and interviews.

Study 2. The second study assessed the feasibility of performing an evaluation with a child test participant paired with a facilitator versus a second child participant (Van Breukelen 2000). The study compared cooperative evaluation by Monk et al (1994) and co-discovery by Kemp and van Gelderen (1996). The two procedures were also compared with respect to the efficiency (number of problems found per participant) and their validity relating to the sources of bias for the children in the two set-ups. This study showed no significant difference of the two methods.

It is too early to draw methodological conclusions from these two studies. From the results obtained so far, it appears that the usability testing methods conceived for adult test participants, work also well with children. Further, the comparative results obtained have been comparable to those concerning adult test participants. Clearly, the two studies described cover a tiny portion of the questions encompassed by the framework described. Currently, more similar studies are under way, that we hope will inform best practice in usability testing with children.

4 References

1. Acuff, D.S. and Reiher, R.H. (1997) What kids buy and why. The psychology of marketing to kids. The free press, New York.
2. Berk, L. (1997) Child development, Allyn and Bacon, Boston.
3. Beusman, J. (2002) An electronic educational interactive game for children in Artis (In Dutch: Een elektronisch educatief interactief spel voor kinderen in Artis), Unpublished M.Sc. thesis, Delft University of Technology, Faculty of Design, Construction and Production.
4. Breukelen, D. Van. (2000) Usability of products for children (in Dutch: gebruiksvriendelijkheid van producten voor kinderen). Unpublished manuscript, Eindhoven University of Technology.
5. Bruckleitner, W., (1999) The state of children's software evaluation – yesterday, today and in the 21st century, Information Technology in Childhood Education, 211-220.
6. Donker, A. & Markopoulos, P., (2002) A comparison of Think Aloud, Questionnaires, and Interviews for Testing Usability with Children. Proceedings HCI 2002, Elsevier (in press).
7. Druin, A., (1999) The role of children in the design of new technology. HCIL Technical Report No.99-23;<http://www.cs.umd.edu/hcil>.
8. Greig, A. and Taylor, J. (1999) Doing research with children. SAGE publications.

9. Hanna, L., Ridsen, K., & Alexander K.J. (1997). Guidelines for usability testing with children. *Interactions* September + October, 9-14.
10. Hanna, L., Ridsen, K., Czerwinski, M., & Alexander, K. (1999). The role of usability research in designing children's computer products. In. Druin, A., (Eds.) *The Design of Children's Technology*, San Francisco, CA: Morgan Kaufman, 3-26.
11. Jacobsen, N.E., (1999) "Usability Evaluation Methods: The reliability and usage of Cognitive Walkthrough and Usability Test, PhD Thesis, Department of Psychology, University of Copenhagen, Denmark.
12. Kemp, J. A. M. and van Gelderen, T. (1996). Co-discovery exploration: an informal method for iterative design of consumer products. In P. W. Jordan, B. Thomas, B. A. Weerdmeester and I. L. McClelland (Eds.), *Usability Evaluation in Industry* (pp. 139-146). London: Taylor and Francis.
13. Monk, A., Wright, P, Haber, J and Davenport, L. (1993). *Improving your human computer interface: a practical technique*. Prentice Hall.
14. Nielsen, J. & Landauer, T.K. (1993) A mathematical model of finding usability problems. In *Proceedings ACM CHI'93*, New York, ACM Press, p. 206-213.
15. Wartella, E., O'Keefe, B., and Scantlin, R., (2000) *Growing up with interactive media: what we know and what we don't know about the impact of new media on children*.