

# The user testing toolset: a decision support system to aid the evaluation of assistive technology products

Andree Woodcock<sup>a, 1</sup>, Simon Fielden<sup>b</sup> and Richard Bartlett<sup>a</sup>

<sup>a</sup> *Centre of Excellence in Product and Automotive Design, Department of Industrial Design, Coventry School of Art and Design Coventry University, Coventry, UK*

<sup>b</sup> *Health Design Technology Institute, Coventry University, Coventry, UK*

**Abstract.** Developers of assistive technology products need to ensure that their offerings meet the requirements of end users, and that usability issues have been discovered prior to manufacture. This may be difficult for SMEs (Small Medium Enterprises) who may lack the necessary skills and resources required to plan and conduct an evaluation. To assist SMEs in the assistive technology market, a stand alone, decision support system was developed to assist in the planning and evaluation of their products, taking into account the resources available, nature of the product being developed and stage of the design process. The responses given by the designer to 40 multiple choice questions are matched against a database of 42 research methods. The methods achieving the highest score in relation to all questions are displayed as the final output. The paper describes the development of the User Testing Toolset (UTT), including the additional functionality provided to ensure that the evaluation methods were correctly matched and weighted to the responses.

Keywords: research methods, usability testing

## 1. Introduction

This paper describes the development of a computer based system to support SMEs in product evaluation. Although developed for the assistive technology sector, it is expected that the User Testing Toolset (UTT) will benefit a wider group of designers.

The UTT was developed as part of the Technology Strategy Board/Engineering and Physical Sciences Research Council funded, i-DEAL project. The academic partners at Coventry University worked with SMEs at the cutting edge of new product development in the areas of assisted living and telemedicine. The overall aims of the project were to develop generic tools and processes to assist SMEs to ensure that their products meet end user needs.

User evaluation is one part of the process, and is very demanding in this sector. Typical challenges include a lack of understanding of the methods that

can be employed to conduct evaluations, lack of resources, access to end users and the required expertise to apply evaluation techniques. Additionally design teams are typically transferring technology to a new market (as in the case of telemedicine) which is uncertain about its purchases (as evidenced by the slow up take of assistive technology), and a user groups which has multiple disabilities.

As part of the project the academic partners assisted SMEs in the early evaluation of a number of products. These evaluations led to significant design changes, with at least one product not being taken forward because of feedback from a focus group. Such decisions are vital to get right given the levels of investment required and resources available in SMEs. As the project progressed it was evident that there was a need to formalise and present assistive technology evaluation techniques and methods in a way which was accessible to designers and

---

<sup>1</sup> Corresponding author: A.Woodcock@coventry.ac.uk

manufacturers, who may have little formal knowledge or training in usability assessment, but who recognise the benefits of liaising with end users throughout the design process. This led to the development of a computer based User Testing Toolset (UTT).

## 2. Establishing the need for a Usability Testing Toolset

To determine how evaluation is currently undertaken in SMEs involved in the design of AT products, eight semi structured interviews were conducted. The companies included a hospital based medical technology company, a company designing circuit boards for medical applications, a design consultancy, young company specializing in cognitive technology, design manufacturing company, a world leader in the design of telecare and assistive technology and a gait analysis SME. Product end users included commissioning organizations, monitoring managers, service telecare providers, housing associations, local councils, operators of telecare services, carers and care recipients.

All companies asserted that they were committed to user engagement throughout the design process from pre brief through to concept and prototype design, after sales follow ups, and cited instances of this. However, there was a reliance on informal qualitative methods (such as informal interviews, user forums and focus groups) and steerage from after sales support teams. Worryingly in some cases evaluation was just with representatives of end users (rather than actual end users), with no formal recording made of an evaluation. Problems associated with the involvement of users included:

1. Finding and accessing representative end users with certain medical conditions.
2. Managing end users expectations, especially in the early stages of concept development '*Users have to understand that the prototypes may not work,*'
3. Relationship management and education - building trust with end users, so that the right level of information is provided, design limitations understood and users trust designers not to 'steal ideas'
4. Costs of running evaluation and management of user groups
5. Perception of the value of end user contributions '*I can't think of an innovation that came direct from an end user*', or '*Well that's pretty difficult;*

*we are dealing with elderly and disabled section of the population so some of them are extremely bright, but the vast majority are being looked after for one reason or another and the very fact that they're being looked after implies that they wouldn't necessarily contribute a great deal to the design process. When you're dealing with people who are already not in the best state of health then that type of feedback is not really forthcoming.'*

The interviews showed that SMEs appreciated the needs for continual user involvement, engaged with different categories of end users, evaluated a wide range of factors than just aesthetics and wanted to build relationships and dialogues with end users and clients. However they relied on a few, qualitative evaluation methods, did not necessarily collect, record or analyze the data, and found it sometimes hard to reach the end users they needed. Given that end users are a valuable resource, it is beholden on evaluators to maximize the quality of the end user experience and ensure that any end user involvement generates the maximum returns.

There may be a gulf between methods used by SMEs and those developed by researchers. The SMEs believe that their current methods were sufficient, so introducing new ones may meet with resistance. However, alternative methods may provide richer material and opportunities for greater engagement. Therefore, there is a need to both inform SMEs about other methods and to provide guidelines on how to design user experiences that are valid, robust and reliable.

## 3. Development of the User Testing Toolset

UTT is a support system to assist designers (especially of Assistive Technology products) in the selection of the most appropriate evaluation method for their given circumstance, taking into factors such as user characteristics, stage of the design process, resource availability, context of use. The software consists of four parts:

1. General advice on how to conduct evaluation.
2. A list of accredited evaluation service providers if more formal evaluation is needed.
3. A brief description of the research methods included in the dataset, and links to more information.
4. Question and answer interface, which takes designers through the factors they should con-



attractive to design students and their lecturers, especially since the system can be used to support the design of any product and at any stage of the design process. The system runs as an independent executable, the questions take about 15 minutes to complete and the output, in terms of the most appropriate research methods is provided immediately.

Factors which needed to be considered in the design of evaluation studies were collated from the literature [e.g. 2, 3 and 4] and the experiences of the project team. This resulted in a set of general questions relating to stages of the development process, purpose of the evaluation, resource availability, end user accessibility, form of the product to be evaluated. Additional questions focused on disability, tasks being supported (e.g. communication, movement), end user group details. Sets of potential answers were developed for each question, and their relevance checked by the project team. For example: the question relating to the length of time available to conduct the evaluation might have the answer set: one day, week, month, three months, longer.

A further review produced a set of over 40 candidate research methods which could potentially be used to evaluate assistive technology products. Experienced researchers mapped the suitability of each research method on to each response. The importance of each group of questions (e.g. user characteristics, stage of the design process) was weighted along with the relative importance of the question in that category. In some cases a particular answer may mean the exclusion of a research method altogether. For example, if the product being developed relates to personal hygiene it may not be appropriate to use an observational study.

#### 4. Behind the system

The UTT is a standalone application, programmed using C-sharp in Microsoft Visual Studio. Its development occurred through the close working of the programmer with a human factor's expert. An iterative process allowed rapid changes and developments, e.g. enabling the testing of different weighting algorithms. Major changes in the interface and functionality were decided on as a result of running the program and reflecting on the system performance.

The heart of UTT is a series of over 40 matrices. One matrix has been developed for each question. In the matrix each research method is weighted against each possible response on a score of -3 to +3 (with an

added value of -999 if the method should never be used). This can lead to matrices of over 400 cells (given that the multiple choice questions can have between 3 and 14 responses).

To assist in the completion of the matrices – cut and paste, scrolling and insertion of standard values across cells were enabled. Initial cell values were provided from the answers of the usability experts who were asked to rate the appropriateness of the research methods to a particular question. Woodcock assessed the answers, and used these to form executive decisions based on her experience of usability testing. Additionally certain questions were weighted more highly than others and adjustments were made for multiple choice answers. Completing the matrices occurred over a 6 month period, with functionality added to make the task easier, and continual testing to ensure that the algorithms were producing 'sensible' answers (see 5.1)..

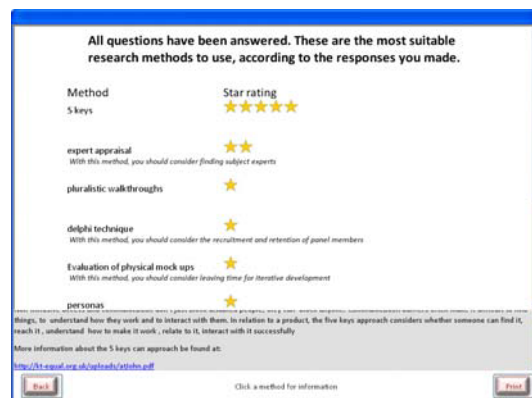


Figure 3. Results screen, showing the most appropriate methods

#### 5. Evaluation of the Toolset

##### 5.1. Internal checks

Given the amount of data which was included in the matrices, internal checks had to be included prior to release to ensure that the weightings given on completion of the toolkit were reliable. The developer's version of the toolset allowed different scoring algorithms and data scores to be adjusted and compared. For example, the effects of increasing the weighting of a particular question, or class of questions could be shown, or the way in which a particular research method had been scored on each ques-

tion. Additionally the numeric values for all 42 methods for any completed session could be shown. This allowed the star rating system to be developed.

### 5.2. External review

The questions and responses were assessed twice by the project team. In attempting to be comprehensive and precise, the language used in some of the questions alienated the end user designers e.g. when using Maslow's hierarchy of needs to categorize the product type. As the system aims to be applicable to a wide variety of products/services/systems and devices it is difficult to find the right terminology to cover all instances (for example the term product, solution, and system have all had their advocates, and none satisfy everyone). In some cases questions seem repetitive, even though they tease out slightly different factors. This can be confusing and frustrating for designers. A second review led to the rephrasing of many questions.

Version 1 of the UTT was evaluated to confirm the appropriateness of the methods when compared with those actually used in evaluation studies. A good match between the suggestions derived from UTT and those used in the evaluations led to confidence in the weighting scales.

A usability study was conducted with 7 representative SMEs to identify usability problems and suggest improvements. The SMEs were engaged in developing discrete AT products - a walking stick, independent living aid, exercise equipment for disabled users, health monitoring devices, mobility aids for children, environmental controls.

Participants described the product they were thinking of evaluating and then worked through the system. Walkthroughs [5] provide an indication of how easily users can perform tasks after little or no training. Observational studies can identify mismatches between the way in which designers think a system should work and users' actual experiences. Figure 2 shows the Q/A interface developed as a result of user feedback.

Participants were asked to comment on the usability of the toolset as they used it. Such verbal commentaries [6] reflect what participants are thinking and can help in understanding user interactions [7]. Instances were identified where participants experienced difficulties operating the toolset. Breakdown analysis [8] was used to rapidly identify and classify problems such as mismatches in how tasks should be

carried out, misunderstandings of the terms used, layout of information and type of feedback.

All participants understood the system, and in most cases felt that it had generated appropriate research methods. However the terminology on many of the questions was still judged to be unfamiliar or inappropriate. Suggestions were given and where possible wording changed, additional user help and explanations were provided. Suggestions were made for optimizing the position of buttons, including more visual feedback to prompt user actions and provide a progression toolbar. Although no comments were provided about the need for system help, SMEs would need help defining their user groups, setting up an evaluation study and understanding different research methods.

The prototype system kept the cumulative scores on the screen while questions were being answered, so the effects of answering a question could be seen on the ordering of the research methods. It was felt that this might be of interest to SMEs and promote learning. It was not. Participants were only interested in the final scores and felt that the provision of superfluous information cluttered the interface and biased their answers. The SMEs requested additional functionality in which more detailed information was provided for the evaluation, such as the questions which should be asked. However, this is not possible as evaluations should be tailored specifically to the product and the users in question.

## 6. Reflections and future work

The interviews with designers showed that they had a considerable amount of interest in engaging users during the development process. However, there is an underlying problem in that the level of input provided by users may be of limited value to designers. A more rigorous approach to conducting evaluation and the use of more appropriate evaluation methods may alleviate this.

The evaluation methods included in the toolset feature those which may be used from initial ideas and pre-concept stages and which allow for more extensive interaction with users, such as codesign and ethnographic methods. Evaluation has to start with an understanding of the product users, and initial ideas evaluated against that knowledge,

Using more immersive methods may help to reduce the barriers between designers and users. The incorporation of such methods into the design

process also necessitates a rethinking of iterative design and evaluation as users become more integrated and integral to the design. This approach is especially important in the design of bespoke assistive technology products. The toolset also acknowledges in its weightings that evaluation does not and should not be limited to laboratory based studies, which fail to recognize the context of use and meaning an assistive technology will have.

### Acknowledgements

The research was supported by grants from the Technology Strategy Board and the Engineering and Physical Science Research Council.

The toolset may be downloaded from the Health Design Technology website, or on application to the corresponding author, A.Woodcock@coventry.ac.uk

### References

- [1] W. Karwowski, (Ed.) *International Encyclopedia of Ergonomics and Human Factors*, 2006 (2nd Edition, Taylor & Francis)
- [2] J. Lumsden, (Ed.) *Handbook of Research on User Interface Design and Evaluation for Mobile Technology*. 2008 (Idea Group, Inc)
- [3] N. Baber and C. Stanton, Factors affecting the selection of methods and techniques prior to conducting a usability evaluation, in P.W. Jordan (ed.), *Usability Evaluation in Industry*, 1996 (Taylor and Francis, London)
- [4] V. Roto, M. Obrist, K. Kaisa Väänänen-Vainio-Mattila, *User Experience Evaluation Methods in Academic and Industrial Contexts*, downloadable from <http://wiki.research.nokia.com/images/9/9a/RotoObristVVM-CHISIGmethods.pdf> accessed 17<sup>th</sup> Sept 2009
- [5] P. G. Polson, C. Lewis, J. Rieman and C. Wharton, *Cognitive Walkthroughs: A Method for Theory-Based Evaluation of User Interfaces*. *International Journal of Man-Machine Studies* 36, 1992, 741-773
- [6] K.A. Ericsson and H.A. Simon, *Protocol Analysis: Verbal Reports as Data*. 1993 (Cambridge MA: MIT Press)
- [7] M. D. Byrne, B. E. John, N.S. Wehrle, and D.C.Crow, *The Tangled Web we Weave: A Taskonomy of www Use*. CHI'99, 1999 (ACM Press)
- [8] A. Woodcock, and S. A. R. Scrivener, 'Breakdown Analysis'. In *Contemporary Ergonomics*. ed. by McCabe P. 2003, (Taylor and Francis; UK)
- [9] A. Woodcock, A. and M. Galer Flyte, Supporting the integration of ergonomics in an engineering design environment, *Tools and Methods for Concurrent Engineering*, 1998, 21-23rd April, Manchester, England, 152-16