The ABCD of Usability Testing

Paula Forbes
Abertay University
Bell St, Dundee
p.forbes@abertay.ac.uk

Stefano de Paoli
Abertay University
Bell St, Dundee
s.dePaoli@abertay.ac.uk

Ken Scott-Brown
Abertay University
Bell St, Dundee
k.scott-brown@abertay.ac.uk

We introduce a methodology for tracking and auditing feedback, errors and suggestions for software packages. This short paper describes how we innovate on the evaluation mechanism, introducing an (Antecedent, Barrier, Consequence and Development) ABCD form, embedded within an eParticipation platform to enable end users to easily report on any usability issues. This methodology will be utilised to improve the STEP cloud e-Participation platform (part of the current STEP Horizon2020 project http://step4youth.eu . The platform is currently being piloted in real life contexts, with the participation of public authorities that are integrating the eParticipation platform into their regular decision-making practices. The project is involving young people, through engagement and motivation strategies and giving them a voice in Environmental decision making at the local level. The pilot evaluation aims to demonstrate how open engagement needs to be embedded within public sector processes and the usability methodology reported here will help to identify the key barriers for wide scale deployment of the platform.

Usability, Feedback, Design, HCI

1. INTRODUCTION

There is widespread problem in the development of software, screen media and large-scale public engagement projects; the problem is how to get feedback from all user groups and stakeholders to the developers quickly, unobtrusively and effectively (Albert, Tullis & Tedesco, 2009). Inevitably, in such projects, the delivery time window is constrained by the stakeholders, and the underlying aims of the project. In addition, the budget is always going to be constrained, and pressure to increase the volume of content, or the scope of new functions can often suck in additional budget (Hawk and Dos Santos, 1991), often leaving the usability budget in deficit before it has even been called into question (Tullis & Albert, 2013). In addition, if a system is being tested in the presence of an evaluator, then there can be an unconscious bias present in the testers to rate any functions positively. This tendency for people to want to give positive feedback is well documented in HCI and Experimental Psychology (Nichols & Maner, 2008), particularly if the participants themselves are pro-social in their outlook and may succumb to the Hawthorne Effect of trying much harder at the task because they are conscious of being observed (e.g. Franke & Kaul, 1978).

The focus on getting feedback in real-time in the company of a moderator can also break up the 'flow' of the user experience. Even, when an experience is self-paced, if the user has to stop and write down every issue they come across as they arise then it can break up the experience. Testers (Users, Developers and Stakeholders) would often rather just get their hands on the system and try it out. If testing is restricted to a single episode, then any emerging idea that happens after the interview or test, is lost, since the time limited questionnaire does not provide an 'incubation' period (e.g. Dorfman, Shames and Kihlstrom, 1996). As the time and money available for user testing with an evaluator is usually low, and the technology now exists to carry out remote testing, and to gain insights from users by asking for feedback, it is important to fully support the feedback that can be provided by users on ad ad hoc basis. Ideally the feedback system should enable users/testers to reflect on their experience and then provide suggestions for improvements.

Clearly longer time-frame testing, and a paradigm that allows interaction and reflection would increase the chance of useful feedback and insight reaching the team of developers. Finally, however, any insight generated is only of use if it is fed back to the team in a timely manner (McKeen, Guimaraes & Wetherbe, 1994). The structure of some experiment-style usability tests can inhibit rapid feedback, since it necessarily involves a temporal disconnect between construction of the system and its testing. Whenever there is a delay for collection, collation and analysis of data, insight feedback can be lost. The problem then is to find a way to share the insights from all partners quickly and efficiently.

For developers, there can be enormous time pressures to fix ongoing issues, having a remote, reflective, asynchronous review based feedback system from users would be helpful, but crucially the feedback received from users should be of a

quality that enables them to understand the issue in order to fix it.

1.1 The STEP approach to Usability Feedback.

The STEP project is developing an online platform (https://en.step.green) with web / social media mining, gamification, machine translation, and visualisation features to engage young people in the decision making process on environmental issues. In order for this to happen the platform itself must be both easy to use and appealing to young people. The project is now in the Pilot phase where it is being tested in 6 different European municipalities / regions.

The novelty of the evaluation approach for the STEP project is the inclusion of an online feedback log to enable users and stakeholders to record information about user experiences. It is set up to allow for live observation and retrospective recall of specific incidents. In particular, it is set up to be as non-intrusive as possible. This guided-defect reporting mechanism helps users through the process, supporting them to provide clear, unbiased reports with rich contextual detail (including the option to provide screenshot images, which should help to provide the developers and the project team with much better feedback to improve the platform. By allowing users to record in an open prompt format the details of their experience, the system avoids some of the 'leading question' problems (Gabbert et al., 2010) associated with structured questionnaires and investigative interviews. Making the responses self-administered allows the participant to record the events at a time and place that suits them and allows them to use their autobiographical memory to respond.

2. ABC(D): ANTECEDENT, 'BEHAVIOUR', CONSEQUENCE (& DEVELOPMENT).

We have developed an Antecedent, Barrier, Consequence and Development (ABCD) report form is intended to supplement other formal evaluations of user experience. The 'ABC' method is derived originally from Clinical Psychology to allow for reflective and unobtrusive collection of incident data retrospectively, but in a structured way, to allow for insight development (Kamps, 2002; Pratt & Dubie, 2008). The overall approach to 'Functional Behaviour Assessment' (Toogood and Timlin, 1996) in Clinical Psychology is to identify the critical *behaviour* (B), then reflect on what the *antecedent* (A) to the behaviour was, and finally record the *consequences* (C) to the behaviour. The practicing clinical psychologist will record these (or ask an observer, or carer, to record these) and then review these later to spot emerging behavioural patterns and to identify solutions.

The ABCD forms purpose is to allow stakeholders a chance to record the experiences and issues that arise when using the STEP platform in a way that allows for their subsequent review, analysis and documentation. Crucial to the STEP project is the facility for the testing and validation to take place in real life contexts. By creating a retrospective log, users can immerse themselves in the actual system and complete a user journey from start to finish, but be enabled, prompted and empowered to record their experiences and insights. Importantly, users will not be presented with the 'Anteceden/Barrier/Consequence/Development' text, they will just be asked a series of short questions in order for them to provide more useful feedback. Questions will be very simple: ie: 'What were you doing?' 'What was the problem?' What effect did this have?' 'What do you think could fix this?'

In the current project we have modified the ABC system to create the ABCD (Antecedent, Barrier, Consequence, *Development*) chart. Examples of ABCD feedback are shown in Figure 1.

What was the 'A'? The 'Antecedent' context.

- What were you doing?

What was the 'B'? The problem 'Barrier'?

- What was the problem?

What was the 'C'? The 'Consequence' for the user?

- What effect did this have?

What is the 'D'? What Development or decision is needed?

- What do you think could fix this?

The ABCD form is for users of the STEP Platform to record information about user experiences using a structured paradigm to enhance the usability of the platform being developed. The form can be used in any AGILE project management setting since they allow for the insertion or inclusion of a suggested remedy. However, this final category is not an obligatory response on the part of the user, often it will be necessary for the developer to provide the development decision in order to overcome the issue being reported.

If time and facilities permit, then the ABCD feedback mechanism can also be incorporated into more traditional user testing. The evaluator can record the whole testing episode with video screen capture, and then carry out an independent post-hoc review to examine the user experience, with the user explaining what they did in a think aloud protocol.overcome the issue being reported. By using video 'logs' in this way, it maximizes the

potential for insight generation since it promotes reflective and recollective processes within a systematic and complete record of activity. By replaying a work episode, the memory component of the task is eliminated, giving time for reflective insight (an 'aha!' moment for example).By pooling all of these data, there is a capability for meta-reflection, by synthesizing all of the ABCD material and reviews of commentaries in a feedback loop.

Antecedent	Behaviour	Consequence What happened as a result?	Decision	Reported by
Trying to register to join a dialogue	Refused to accept the postcode	Could not register, could not participate in the dialogue	Allow alphabetic data range for postcode in addition to numerical	
Setting up a group of people to invite to new challenge	Unintuitive method for adding new people (not already signed up) caused a problem	Could not proceed without seeking assistance. Resorted to inviting each person individually	Allow new users to be added to a group and invited at a later time (drag & drop for group creation?)	Paula Forbes

Figure 1: Example of ABCD report.

3 MISMATCH BETWEEN USER REPORTING AND DEVELOPER NEEDS

Reporting usability defects can be challenging, specifically getting the developers to understand the issue and to appreciate its importance is often a major challenge. Yusop et al. (2016) surveyed both open source and industrial software developers about their usability defect reporting practices and found that reporting the 'cause' of the problem is the most difficult information to provide whereas software developers consider this to be the most helpful information in enabling them to fix the issue. Getting user feedback to developers in an appropriate format can be problematic, the developers own issue tracking mechanisms (such as Trac & JIRA) are set up very much for the developers' own use and not are intended for end-users to report issues. The form fields are confusing for non-developers who are unlikely to know what many of the fields are asking for (for example, Sprint cycle, Linked issues, MoSCoW analysis) or to know how to prioritise the issue they are reporting.

Developers may also have a tendency to view usability issues as 'nitpicking' and of less importance than other types of software issues (Yusop & Vasa, 2016). One of the authors has previous experience of working on usability issues in scientific software development (Sloan *et al.* 2009). The experience of the scientific software development project highlighted that to integrate usability and UCD into the software platform, effective communication between end users, the usability team and the developers was vital. Each week testing was carried out with end users and issues were recorded (with details provided in a

similar way to the ABCD report although no suggested 'fix' was listed). The reports were then shown to the Developers who then added their own column at the end (in essence, a Development or 'D' field) for a suggested fix labelled 'ticket'. These tickets were then transferred by the Developers, using their own language and terminology, into their 'Trac®' Issue tracker system (Macaulay et al. 2009). This approach avoided some of the problems of the misunderstanding of reported issues described by Yusop & Vasa (2016), reducing the gap between end user needs and the needs of the software developer to understand the issue. By taking into account this approach, the aim is to both reduce the barriers for end user issue reporting, and also to try and lead the user into making more useful and meaningful reports that are more easily assimilated by the software developers.

4 CONCLUSIONS

In order to produce more useful usability feedback and taking advice from Kujala (2003) that user tests must be of lower cost to implement, we present a method that can be flexibly deployed in an iterative design project to gain insights and spot usability problems for operations in the field. The approach is consistent with recent reviews of usability testing that emphasise the benefits of qualitative as well as quantitative data obtained from transcripts of reviews (Ebling and John, 2000). Using the ABCD form in an online format allows for optimal user feedback to be provided over time, and also ensures that data from usability testing can be incorporated into the developers own systems in a meaningful way.

The ABCD chart can be flexibly incorporated into existing empirical usability techniques and methods such as Cognitive Walkthroughs (e.g. Polsom, Lewis, Rieman, and Wharton, 1992); Heuristic Evalution (Nielsen, J., and Molich, 1990; Nielsen, 1992; Neilsen and Phillips, 1993). Cognitive walkthroughs and cued retrospective commentaries are planned as part of the STEP project evaluation, and these, in combination with post-study interviews and surveys, should be effective in capturing usability and other issues with the platform. Previous studies have shown that combining appropriate usability methods is much more effective than a single method (Walji et al. 2016; Middleton et al. 2013). The User testing method is best for detecting specific performance problems as they can be reported as they are encountered. Survey and interview methods can help to verify reoccurring usability problems, but are of limited value used in isolation. Walji et al. (2016) concluded that no single method was successful at capturing all usability problems, therefore a combination of different techniques that

complement one another is necessary to adequately evaluate the system.

In summary, the online ABCD form provides a mechanism to support remote user feedback, and a means for reflection to be incorporated into the feedback process by facilitating the user to document the retrospective recording of insights. Unlike experimenter led usability sessions, this method provides thinking space after a user experience to provide feedback. Harnessing autobiographical and episodic memory also creates an opportunity for users to 'join up' insights from multiple instances.

ACKNOWLEDGMENTS

REFERENCES

Albert, W., & Tullis, T. (2013). *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. 2nd Ed. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA ©2013

Albert, W., Tullis, T., & Tedesco, D. (2009). *Beyond the usability lab: Conducting large-scale online user experience studies*. Morgan Kaufmann.

Dorfman, J., Shames, V. A., & Kihlstrom, J. F. (1996). Intuition, incubation, and insight: Implicit cognition in problem solving. In Underwood, Geoffrey D. M. (Ed). (1996). Implicit cognition (pp. 257-296). New York, NY, US: Oxford University Press, 305 pp.

Ebling, M. R. and John, B. E. (2000). On the contributions of different empirical data in usability testing. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques* (DIS '00), Daniel Boyarski and Wendy A. Kellogg (Eds.). ACM, New York, NY, USA, 289-296.

Franke R.H., & Kaul, J. D. (1978) The Hawthorne experiments: First statistical interpretation. Am Sociol Rev. 43:623–643. doi: 10.2307/2094540.

Gabbert, F., Hope, L., & Fisher, R. P. (2009). Protecting eyewitness evidence: Examining the efficacy of a self-administered interview tool. *Law and human behavior*, *33*(4), 298-307.

Hawk, S. R. and Dos Santos, B. L. 1991, Successful system development: The effect of situational factors on alternate user roles. IEEE Transactions on Engineering Management, 38 4), 316 ± 327.

Kamps, D. M. (2002). Preventing problems in improving behavior. In B. Algozzine & P. Kay (Eds.), *Preventing problem behavior* (pp. 11–36). Thousands Oaks, CA: Corwin Press.

Kujala, S. (2003). User involvement: a review of the benefits and challenges. *Behaviour & information technology*, *22*(1), 1-16.

Macaulay, C., Sloan, D., Jiang, X., Forbes, P., Loynton, S., Swedlow, J. R., & Gregor, P. (2009). Usability and user-centered design in scientific software development. *IEEE Software*, *26*(1), 96.

McKeen, J. D., Guimaraes, T., & Wetherbe, J. C. (1994). The relationship between user participation and user satisfaction: an investigation of four contingency factors. *Mis Quarterly*, 427-451.

Nichols, A. L., & Maner, J. K. (2008). The good subject effect: Investigating participant demand characteristics. Journal of General Psychology, 135, 151-165.

Nielsen, J. 1994. Usability inspection methods. In *Conference Companion on Human Factors in Computing Systems* (CHI '94), Catherine Plaisant (Ed.). ACM, New York, NY, USA, 413-414.

Nielsen, J., and Molich, R. (1990). Heuristic evaluation of user interfaces. Proc. ACM CHI'90 (Seattle, WA, 1–5 April), 249–256.

Nielsen, J. (1992). Finding usability problems through heuristic evaluation, *Proceedings of the Conference on Human Factors in Computing System (CHI 92)*, Monterey, CA, May 3-7, 1992, 373-380.

Nielsen, J., and Phillips, V, (1993). Estimating the relative usability of two interfaces: heuristic, formal, and empirical methods compared, *Proceedings of the Conference on Human Factors in Computing System (CHI 93)*, Amsterdam, April 24-29, 1993, 214-221.

Polsom P., Lewis, C., Rieman, J., and Wharton, C. (1992). Cognitive walkthroughs: A method for theory-based evaluation of user interfaces. International Journal of Man- Machine Studies, 36,5, 741–773.

Pratt, C., & Dubie, M. (2008). Observing behavior using a-b-c data. The Reporter, 14(1), 1-4.

Sloan, D., Macaulay, C., Forbes, P., & Loynton, S. (2009, September). User research in a scientific software development project. In *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology* (pp. 423-429). British Computer Society.

Toogood, S. and Timlin, K., 1996. The Functional Assessment of Challenging Behaviour: A Comparison of Informant-based, Experimental and Descriptive Methods. *Journal of Applied Research in Intellectual Disabilities*, *9*(3), pp.206-222.