

Improving Children's Self-Report in User-Centered Evaluations

Marco Pasch

University of Lugano
Via Buffi 13, 6900 Lugano,
Switzerland
marco.pasch@usi.ch

ABSTRACT

The envisaged contribution of the research presented in this paper is to improve users' self-report, a popular way to assess user experiences. The starting point of the research is an assessment of the previously presented Sensual Evaluation Instrument objects when used by children. Our assessment is set in a naturalistic setting, i.e. a primary school in southern Switzerland. Should the assessment be positive, we plan to explore further ways of using physical objects to support self-report. This should result in establishing a more robust version of self-report from children, which up to now suffers from being dependent on facilitators and requiring children to possess high levels of reflection and linguistic skills.

Categories and Subject Descriptors

H.5.1 [INFORMATION INTERFACES AND PRESENTATION]: Evaluation/methodology

General Terms

Human Factors

Keywords

Self-report, evaluation, children

INTRODUCTION

A number of methods have been proposed to assess users' subjective experiences when engaging with interactive applications. These range from self-report methods to measuring physiological data during interaction. While physiological data can objectively show arousal states, it fails in capturing how users subjectively appraise their experiences. Though self-report can give us the subjective

appraisal, in practice it suffers from requiring users to have excellent linguistic skills and demanding a certain level of reflection in the first place. In particular for children these drawbacks have been addressed by employing facilitators that assist children, trying to keep them engaged and to elicit feedback. Yet, their presence can lead to an increased experimenter effect. My research aims to improve measures of self-report, based on an embodied view to interaction.

USER EXPERIENCE ASSESSMENT INVENTORY

Before stating the research problem further, I want to provide an (certainly non-exhaustive) overview of existing methods aimed at assessing the User eXperience (UX). This provides the foundation for later grounding our own approach into the body of previous research.

Methods for assessing the user experience can be first and foremost distinguished by whether they are based on subjective self-report of users or obtained from objective measurements of users using an interactive application.

Objective measurements include data from observation of the users' behavior, physiological data, and voice. Behavior can be assessed by e.g. measuring the performance of the user in terms of speed required for task-solving, measuring how often and with how much force the user clicks the mouse or punches keys, or by measuring the eye gaze of the user. In particular eye gaze is a popular measure for getting an idea which items catch the attention of the user. Physiological data is popular in the field of affective computing. Measuring heart rate, skin conductance, and other items allows inferences on arousal states of users and whether their valence is positive or negative. Often several measures are combined and data fusion algorithms are applied to get a better assessment of users' affective states.

The subjective user experience can be obtained through verbal and non-verbal methods. In practice, verbal methods such as post-experimental interviews and questionnaires are used widely. A more recent development are non-verbal methods. Lottridge (2009) uses continuous sliders that users are instructed to use during interaction, like Likert scales

ranging from very positive to very negative. Xu (2009) has proposed to let test participants – in this case children – make free drawings in which they express their experience and to have the drawings reviewed by experts.

RESEARCH PROBLEM

My research starts from identifying shortcomings of the aforementioned methods. Although objective measurements come with the promise of providing an objective picture of users' affective states, they suffer from not being able to grasp how the user individually assesses the interaction. Gaver (2009) provides an excellent example to clarify this point: He describes going fishing on an afternoon and upon making a big catch feeling very happy. He points out that this would have been easily identifiable by measuring e.g. physiological data. He then describes playing with his kids in the living room later that day. Again, the fact that he was happy is supposed to be easily identifiable. Gaver's main point is that although an automated emotion recognition system would assign the label "happy" to both situations, for Gaver himself the situations were very different. Although he did indeed feel happy in both situations, he would never put them on par. The mere assignment of a label misses so much of the actual experience. So if we really want to understand the experiences of users, we cannot avoid asking them for their individual assessments.

But also self-report measures suffer from shortcomings. Interviews require users to have excellent linguistic skills to give an account of their experiences. In particular for children these drawbacks have been addressed by employing facilitators that assist children, trying to keep them engaged and to elicit feedback. Yet, facilitators can influence children in these evaluations in various ways, which leads to an increased experimenter effect.

Table 1: Subjective and objective measures vs. time of measurement

	Subjective Self-Report	Objective Measurements
During Interaction	Think Aloud Emotion Slider	Behaviour Physiological Data Gaze Voice Prosody
After Interaction	Interviews Questionnaires Drawings	

Another important aspect is whether the method captures what happens during interaction or just a post-experiment assessment. Table 1 shows the methods introduced so far and differentiates whether they capture the experience during or after an interaction. As can be seen, while all objective measurements are done during an interaction, self-report methods differ in this respect. Interviews, questionnaires and drawings happen after an interaction. This comes with the disadvantage that what we actually get

from the users here is a mere summary of the actual experience. From these summaries it is hard to go back to specific events during the interaction and the role that this events played for the remainder of the interaction.

The established Think Aloud method, which makes users speak out loud what they are thinking and doing during an interaction obviously takes place during interaction. The downside of this method is that the constant verbalization of thought hinders a deeper immersion into the interaction and thus heavily influences what it is actually supposed to measure. Lottridge's (2009) sliders also measure self-report during interaction, but suffer from the reduction of user feedback to one or several dimensions (arousal, valence, etc.). Again, the uniqueness of an experience is lost.

AIMS AND APPROACH

My research aims at establishing a further channel of affective feedback by employing physical objects that users interact with at evaluation time. Apart from targeting adults, I want to find better ways to assess children's user experience, as existing methods are often not suited for them. In particular for children there are clear advances to an embodied, nonverbal way of eliciting cues about the affective state. Antle et al. (2009) have shown that children are able to act on preconscious knowledge, though they are not (yet) able to verbalize it and put forward that the difference between preconscious and explicit knowledge is bigger in children than in adults (Antle et al., 2008). My hypothesis is that there is a parallel for affective states and that we can learn a lot about a child's affective state through cues from an embodied interaction. These cues can then either be evaluated independently or used to get better feedback in post-experiment interviews or similar self-report measures.

A first step in the project was an extensive literature review on existing methods assessing the user experience. This critical assessment resulted in two position papers: The first paper (Pasch and Landoni, 2009) discussed short-comings of existing evaluation methods when used for children and explored the approach of using physical objects to facilitate reflection and verbalization of emotional states during interaction. The second paper (Pasch and Landoni, 2010) focuses on video games as emotionally rich environments and how games can be a fruitful testbed for our approach.

In a second step I am evaluating the use of the Sensual Evaluation Instrument (SEI) objects for our purposes. The objects consist of 8 hand-sized sculptures developed by Isbister et al. (2006). The objects all share the same color and texture, and only differ in their shapes, which range from pointy edges to bulbous curves. Figure 1 shows the SEI objects.



Fig.1: Sensual Evaluation Instrument (SEI) objects

Currently, we are running a series of observations and consecutive experiments involving children from a local primary school in Lugano, Switzerland. We have adapted the original validation study by Isbister et al. (2006) involving a small group of HCI researchers in laboratory setting, to run with children (age 6-11) in a more formal but naturalistic educational scenario. We have devised a series of experiments to fit into 4 different educational scenarios:

1. Primary four (9-10 years old) children taking their English and ICT class at the same time: this is the most formal scenario and its purpose is to test SEI in a formal educational setting where children engage with ICT under teacher supervision and with well defined educational objectives.
2. A mixed group of primary school children attending after-school. This is the most playful scenario where children engage in their after-school time with video games, such as Wii Sports on the Nintendo Wii platform.
3. Primary one and three children taking their physical education class. This is another formal but playful setting that has emerged from our conversations with teachers as we realized that one of the explicit learning objectives of physical education is to teach children how to deal with positive and negative emotions making it an ideal scenario for our study. This will combine the educational with the playful setting enabling us to work within well-defined learning objectives.
4. Primary four children taking part in a cinelogy session. This is a one off session during which children are shown an emotion-charged movie and then invited to discuss their feelings with a psychologist running the session.

Once parents and teachers were briefed about the research in terms of objectives and implications on teaching and learning, we started experiments with the first two groups of children in October 2009. This meant for the first roughly four weeks to sit in the English and ICT class and observe children's behavior with each other, the teacher,

and with computers. From this we could learn the class procedures and plan better how to integrate the SEI objects into class. Maybe more importantly it also gave the children time to get used to us, so that our presence became something normal for them. Blending in is important as it reduces the effect of the presence of experimenters in natural settings. In addition, we brought a Nintendo Wii to the after-school program once a week.

We then started out doing ad-hoc interviews with the children after class, resp. after they played with the Wii in after-school. Varying interviewing them alone or in groups, we got an idea of what kind of feedback we can expect from them. This is the state of the experiments at the time of writing this report. In the upcoming weeks we will change the interview style to more thorough interviews, aiming at more tangible results. The plan is to have more structured interviews and to transcribe them. With the transcripts we can then e.g. analyze, whether the children show bigger active vocabulary after having used the objects, as compared to when they are just interviewed.

INTERMEDIARY RESULTS AND FUTURE DEVELOPMENT

Early outcomes of the ad-hoc interviews point on some relevant difference in the use of the SEI by children. They regard them more as toys than tools for evaluation and interpret shapes in a more amenable and context sensitive fashion. In fact, each of the children involved has volunteered their favorite shape and e.g. they have renamed the "spiky" object with a much softer name: sea-star or sea-urchin. Interesting differences have emerged also between younger (primary 1 and 2, i.e. 6/7 years old) and older (primary 4 and 5, i.e. 9/11 years old) children and among this second group between girls and boys.

A report describing the findings from these experiments and how these can be interpreted in order to redesign, adapt or even create new tools to achieve the project objectives will be ready as soon as the experiments are completed. Having to adapt to the school calendar has caused a little delay on our expected timetable here.

Our plans for future development are yet depending on the outcomes of our SEI evaluation study. For the time being, we have developed potential lines of research, which are briefly described in the following. One idea is to involve children in the redesign of the SEI objects. Apart from getting their input from interviews and similar evaluation techniques, we want to engage them in participatory design sessions in parallel to our user studies. This includes brainstorm sessions and inviting them to model new sets of SEI objects with modeling clay, plastilina or play-doh.

A further idea is to equip the SEI objects - or a redesigned version of them - with sensors to elicit data about their usage automatically. We are thinking here of accelerometers measuring the amount of movement of an object, tactile sensors measuring the force that users exert

on an object, etc. The application of this data is not to create automated profiles of usage, as this would lead to the shortcomings of automated approaches that we discussed earlier, i.e. in short, failure to capture the uniqueness of the interaction. Instead we want to show this data, i.e. suited visualizations of the data, to users in order to help them reflect on their experience and verbalize them in post-experiment interviews.

CONCLUSION

This research starts from recognizing the shortcomings of traditional techniques when involving children in evaluation experiments. By using physical objects to elicit affective feedback from children, we not only aim at overcoming the need of facilitators, who can bias outcomes but also at gaining access to preconscious, tacit knowledge that children often have, but are not able to verbalize.

The goal is to provide new, effective tools to support evaluators in different scenarios such as formative evaluations involving children and/or systems expected to have a strong emotional impact. While the research currently focuses on children, we aim to establish an evaluation method that is also suited for adults.

Our evaluations are set in a naturalistic setting, i.e. for the moment in a primary school. Here we can evaluate in a familiar environment and within children's social context.

The tools we propose should help to get a clearer picture of the affective states of users and to get more accurate and representative feedback to be gathered by evaluators. This should provide designers with better communication tools to support a fruitful and more natural collaboration with users, which is expected to result in improved system usability.

ACKNOWLEDGMENTS

We thank Kia Höök and the Interaction lab at the Swedish Institute of Computer Science (SICS) for providing us with a set of the SEI objects and their friendly collaboration.

REFERENCES

1. A. N. Antle, G. Corness, and M. Droumeva. What the body knows: Exploring the benefits of embodied metaphors in hybrid physical digital environments. *Interacting with Computers*, 21(1-2):66 – 75, 2009. Special issue: Enactive Interfaces.
2. A. N. Antle, M. Droumeva, and G. Corness. Playing with the sound maker: do embodied metaphors help children learn? In *IDC '08: Proceedings of the 7th international conference on Interaction design and children*, pages 178–185, New York, NY, USA, 2008. ACM.
3. W. Gaver. Designing for emotion (among other things). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3597–3604, 2009.
4. K. Isbister, K. Höök, M. Sharp, and J. Laaksohlahti. The sensual evaluation instrument: developing an affective evaluation tool. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1163–1172, New York, NY, USA, 2006. ACM.
5. D. Lottridge. Evaluating human computer interaction through self-rated emotion. *Human-Computer Interaction – INTERACT 2009*, pages 860–863, 2009.
6. M. Pasch, M. Landoni. Using Physical Objects to Capture Children's Affective States in User-centered Evaluations. Position Paper presented at the IDC'09 workshop: Children and Embodied Interaction, Como, Italy, June 2009.
7. M. Pasch, M. Landoni. Video Games as Research Stimuli to Study New Ways to Assess User Experience. Position Paper accepted for the CHI 2010 Workshop: Video Games as Research Instruments, Atlanta, USA, April 2010.
8. D. Xu, J. C. Read, G. Sim, and B. McManus. Experience it, draw it, rate it: capture children's experiences with their drawings. In *IDC '09: Proceedings of the 8th International Conference on Interaction Design and Children*, pages 266–270, New York, NY, USA, 2009. ACM.