Teachers' involvement in usability testing with children

Sofia Pardo, Frank Vetere and Steve Howard

Department of Information Systems, The University of Melbourne

Level 4 111 Barry St. Carlton, Melbourne, Australia

{miriamp, fv, showard}@ unimelb.edu.au

+ 61 3 83441500

ABSTRACT

This paper describes the involvement of teachers as facilitators in user-based evaluation of educational software. Findings from a case study that compare the behaviour of teachers and designers during the evaluation indicate that there is no strong evidence to support the current practice of teachers' tangential participation in the evaluation of educational software. Further research is needed to better understand the role teachers can play in this type of evaluation so that their knowledge and experience can positively contribute to the evaluation process and outcomes, whilst retaining the children's valuable contribution.

Keywords

User based evaluation, usability, children, teachers, educational software.

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation (e.g. HCI): User Interfaces - Evaluation/Methodology

INTRODUCTION

The benefits of collecting data from children as tester of computer programs have been widely documented [2, 5]. Data from children afford insight into their perspectives as a user group and can supply evidence on the effectiveness of design decisions [5]. However, these benefits are not without their challenges. Children's limited verbal and social skills, and their relative powerlessness are barriers that designers may have to overcome.

The evaluation of educational computer programs places other requirement on the already challenging scenario of involving children in user-based evaluations. Children can not provide feedback on the learning goals they haven't yet attained. Hence the input of other stakeholders, such as teachers, psychologists or educational specialists is often required [8].

Designers are normally provided with guidelines and methods to facilitate their interactions with the children.

However less support is given to the challenge of coordinating the contributions of different stakeholders, designers, teachers and children. As a result, the current practice often places teachers in a peripheral position, so that their input is taken into account as informants in the prior stages of the evaluation without affecting the fragile equation designer-child.

Teachers' input is often sought regarding the evaluation materials, the choice of evaluation tasks, the selection of the children and the allocation of physical resources for the evaluation to take place [1]

On the other hand, the actual evaluation is seen as a unique place created for the children with the purpose of hearing their voices without the interference from teachers and other adults. It is believed that teachers might assume a traditional teaching role and the children will enact a passive one. This assumption has given teachers' participation marginal influence on the evaluation process despite the lack of empirical evidence to support such assumption.

Drawing from Prensky's [7] work, we have argued elsewhere [6] that when approaching technology teachers and children's traditional roles are somehow reversed. Children don't expect the teachers to know everything and teachers assume a learner role. This paper addresses the question of what is the impact the involvement of teachers in user-based evaluation with children has on the evaluation process and outcomes.

METHODOLOGY

We examine the role of teachers within child-based evaluation by reporting on the actual evaluation of different educational software created by a Melbourne company. No changes were made to their existing practice for the purpose of this study. This research compared the behaviour of the teachers and designer's as facilitators in order to identify the impact of these differences on the evaluation process.

Participants

Participants included:

- One designer from a software company.

- Two teachers from two urban primary schools who had previous experience running this type of evaluation.

- Sixteen students from grades 2-4 (7-9 years old) working in pairs who were selected by their teachers according to a criterion set by the software company.

Evaluation Set up

Four different educational computer programs were evaluated. Each computer program was evaluated twice, giving a total of eight evaluations. Each of the eight evaluations consisted of two students and one facilitator. The designer acted as facilitator in four evaluations and two teachers were each facilitators in two other evaluations (see table 1).

	Computer Program			
	A	B	C	D
Teacher 1	x	x		
Teacher 2			x	x
Designer	x	x	x	x

Table 1

Teachers were given a task list, containing a set of questions to ask the students and guide their observations. They were also provided with the computer program a week before the evaluation for purposes of familiarisation. The evaluations took place at the primary schools outside classroom hours and lasted around 45 minutes. Teachers and designers ran the evaluations independently.

Evaluation Materials

The questions on the task list mostly addressed learning issues relating the children's understanding of a concept as well as the processes they go through to achieve that understanding. A smaller number of questions concerned usability issues. The fewer usability questions maybe a consequence of the high familiarity that the design team had with the major interface problems. Finally a reduced number of questions addressed engagement issues. This can be explained by the fact that children's non-verbal behaviour might be a better indicative of the level of engagement than their direct responses. The evaluation materials served to better understand the software company's purpose for the evaluation.

Data Collection

The data collection was conducted at three different stages of the evaluation process:

- *Before the Evaluation.* We interviewed the designer before the evaluation in order to understand the overall purpose of the evaluation and the type of questions that were going to be asked. Interview with the designer was audio recorded and the task lists containing the questions were collected. A coordinator from the software company was present during the interview.

- *During the Evaluation.* Evaluations were video recorded and field notes were taken in order to identify the novel questions, i.e. those asked by the teachers and the designer that were not on the task list provided.

- *After the Evaluation.* Debriefings with the designer and the teachers were audio recorded. Debriefings with students were conducted in a group interview format, four groups of 4 students each video recorded. Debriefings aimed to gather all participants' perceptions of the teachers' and designer's participation in the evaluations.

FINDINGS

The findings cover two topics. The first concerns the differences between the teachers' and the designer's behaviour during the evaluation, and the second concerns the participants' perceptions of the involvement of teachers and designers in the evaluation process.

Comparing Teachers and Designer's Behaviours

In order to understand the behaviour differences between the teachers and the designer in the evaluation processes, attention was paid to the unscripted questions they asked the children that were not on the task list and to their written notes that reflected their perceptions.

The unscripted (or novel) questions and written notes were classified into four categories: Usability, Learning, Both and Other, depending on the issue that was being addressed. It is important to highlight that these categories are not mutually exclusive since usability and learning are meant to interact in a way that facilitates conceptual growth. However, this classification allows a better view of the main issues identified in the evaluation.

- *Usability*: Navigation and interaction with the interface of the computer program.

- *Learning*: Students' understanding of the concept/s that was conveyed through the computer program.

- *Both*: Overlap of learning and usability issues. The majority regarded the use of language, in some cases referred to the concepts the computer program was conveying, while in other cases it referred to expressions the children were not familiar with and hindered the interaction with the interface.

- *Other*: The evaluation process as such, for instance, whether the children were cooperating, sharing the mouse, or being very quiet as they worked through the different games or tasks.

Regarding the novel questions, we expect that teachers, given their training, would focus mainly on the learning issues of the computer program [3]. However, it was found that they tended to ask more novel questions addressing usability issues (Figure 1).
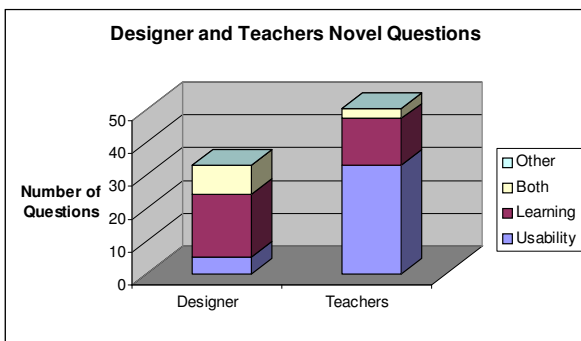
Figure 1: Comparison of Teachers and Designer Novel Questions

This may be explained by the novelty of the computer program. Since teachers were not as familiar with it and because it is not their area of expertise they might have tried to compensate by paying special attention to the children's interactions with the interface. Furthermore, the learning issues might have been well covered by the questions on the task list, so that teachers had enough evidence of the children's conceptual understanding.

Additionally, teachers weren't aware of the learning goals or the rationale behind the computer programs. These were gleaned from the questions on the task list and from the actual use of the prototypes. This could have influenced teachers' tendency to focus more on the usability issues, since these relate more to the visible aspects of the interaction with the computer program and to a lesser extend to the intangible learning goals.

On the other hand, the novel questions asked by the designer were consistent with the evaluation objectives and the questions on the task list. The designer tended to ask more questions regarding learning issues. This tendency can be explained because of the designer familiarity with the computer program which left more room to explore the student learning experience. Additionally, the designer had more limited experience in obtaining evidence of students' conceptual understanding and learning processes through conversation and observation as opposed to the teacher, which suggests an extra effort on trying to gather these types of data.

Regarding teachers and designer written notes, the number and type of problems identified were very similar in almost all of the eight computer programs. It seems that knowing the children beforehand did not make any difference to the evaluation outcomes. Further more, teachers, despite their background and lack of training conducting this type of evaluations, were able to account for both usability and learning issues with the computer program.

Another distinction identified between teachers' and designer's written notes was the teachers' tendency to identify each child by name in their. Although this was recommended on the evaluation materials, the designer considered the children's answers as a whole and his/her

notes reflected more the discussion between the children, and the main arguments. It can be suggested that teachers played the role of data collectors and felt more compelled to gather the answers to the questions set by the designers rather than follow the children's interactions in a more open-ended fashion.

Students, teachers and designer's perceptions

All participants were asked to think about the involvement of teachers and designers in the evaluation with the students.

Students: When children were asked to comment on their perception of teachers and designers involvement they referred to their common experience of having "helpers" from outside the school coming to their classrooms. Teachers were described as the ones that know and can do everything, while the helper was described as a 'small teacher' who is learning and doesn't know the students names as opposed to the teachers.

Children described teachers as the ones who correct their work; however when they were asked to compare the evaluation on the computer with the exams they normally do in their classroom some expressed a different view. The evaluation on the computer was not considered like an exam by most of the children, because it was fun. There were not right or wrong answers and the teachers were not marking their work.

Designer: The designer considered involving teachers was very valuable because of two main reasons. Firstly because teachers are familiar with the children's prior knowledge and whether they are high or low achieving students. And secondly, because it might be reassuring for the students to have someone they know conduct the evaluation.

On the downside, the limited experience teachers have running these evaluations and limited knowledge about the computer program background was seen as an obstacle by the designer when trying to meet the evaluation objectives. It was considered that when teachers have the opportunity of seeing a designer running an evaluation, they might improve their performance and therefore the quality and quantity of the feedback collected.

Teachers: They described themselves as having more 'insight' into the children, as opposed to the designer who has more 'insight' into the software. Through their experiences in the classroom they have learnt how to explore what children think and get them to elaborate on their ideas and thoughts.

Some of the challenges teachers experienced during the evaluation was writing down on paper all the relevant information, and having to juggle their own observations with the children's responses to the questions. Teachers also found difficult to relate the purpose of the computer program to their perception of what the children were learning.

CONCLUSION

The involvement of teachers as facilitators in child-based evaluations of educational software is described at two different levels. Firstly at a process level, on the interactions between the students, the teachers and the designer during the evaluation. And secondly, at an outcome level, on the problems that were identified with the computer programs.

At a process level, the well established concern of the power differential between students and teachers as an obstacle when trying to collect genuine feedback from the students [4] has been shown not to be critical. Students in this research project did acknowledge teachers' authority, but it wasn't experienced as an obstacle or something that would inhibit their participation. Having the teacher observing and asking questions as they worked through the computer program didn't resemble, according to the students, the traditional exams in which there are right and wrong answers.

Regarding the designer participation, students situated the teachers in a super-ordinate position, which can be a result of the teachers playing the liaison role between the designers and the students; so that teachers introduced the evaluation activity and the designer. Likewise the designer was aware of this dynamic. Teachers are the gatekeepers and provided their time and resources for the evaluation to take place. Nevertheless both designer and teachers acknowledge teachers' lack of experience and the need to follow the designer's instructions and advice as close as possible in order to collect the children's feedback as required.

Teachers were aware of the importance of hearing the children's voices and put special effort in avoiding leading questions. It was clear to them that the evaluation required them not to play a teaching role, instead to be a curious observer of the children's interactions with the computer program.

At the outcome level, teachers and designer identified very similar problems with the computer programs. This suggests that with little training teachers are able to collect relevant feedback without being hindered by the fact of being seen as an authority by the students. It can be argued that given the similarity of the evaluation outcomes there are no strong reasons to exclude teachers from the evaluation.

In sum, teachers' behaviour as facilitator was highly shaped by the artifacts that were given to them and by the evaluation set up provided by the designer. This case study has shown that the current practice of placing teachers into a secondary role in this type of evaluation does not have empirical support. It is argued that designers need to develop more sophisticated ways of involving teachers in child-based evaluations so that their contribution can be maximized.

REFERENCES

1. Africano, D., Berg, S., Lundholm, P., Nilbrink, F. and Persson, A. Designing Tangible Interfaces for Children's Collaboration. *CHI 2004*, Viena (2004) ACM Press.

2. Baauw, E. and Markopoulos, P. A comparison of thin-aloud and port-task interview for usability testing with children. *IDC 2004*, June, 1-3, (2004) ACM 115-116.

3. Draper, S. W., Brown, M. I., Henderson, F. P. and McAteer, E. Integrative Evaluation: An emerging role for Classroom Studies of CAL. *Computers in Education*. Vol. 26. No. 1-3 (1996), 17-32.

4. Druin, A. The role of children in the design of new technology. *Behaviour & Information Technology, an international journal on human aspects of computing* 21 (1), (2002) Taylor & Francis.

5. Hanna, L., Risden, K. and Alexander, K.J. Guidelines for Usability Testing with Children. *Interactions (September/October* 1997), 9-14.

6. Pardo, S., Vetere, F. and Howard, S. Broadening Stakeholder Involvement in UCD: Designers' Perspectives on Child-Centred Design. In Proceedings of OZCHI 2005 (Canberra, Australia, November 2005), ACM Press.

7. Prensky, M. Digital Natives, Digital Immigrants. *On The Horizon*, Vol. 9 No5 (2001) NCB University Press.

8. Scaife, M., Rogers, Y., Aldrich, F. and Davies, M. Designing for or Designing with? Informant design for interactive learning environments. *Proceedings of CHI* (1997) ACM, 343-350.