

Experiences with structured interviewing of children during usability tests

Citation for published version (APA):

Vermeeren, A., Bekker, M. M., van Kesteren, I. E. H., & De Ridder, H. (2007). Experiences with structured interviewing of children during usability tests. In *People and Computers XXI HCI : But Not as We Know It - Proceedings of HCI 2007: The 21st British HCI Group Annual Conference* (Vol. 1). British Computer Society (BCS).

Document status and date:

Published: 01/12/2007

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Experiences with Structured Interviewing of Children During Usability Tests

A. Vermeeren¹, M.M. Bekker², I.E.H. van Kesteren³, H. de Ridder¹

¹TU Delft, The Netherlands
Industrial Design
Landbergstraat 15
NL 2628 CE, Delft
+31 15 2784218

²TU Eindhoven, The Netherlands
Industrial Design
Den Dolech 21
NL 5600 MB Eindhoven
+31 40 2475239

³TU Delft, The Netherlands
Design Engineering
Landbergstraat 15
NL 2628 CE, Delft
+31 15 2785730

{a.p.o.s.vermeeren, h.deridder}
@tudelft.nl

m.m.bekker@tue.nl

i.e.h.vankesteren@tudelft.nl

ABSTRACT

In this paper we describe an exploratory study on the use of a structured interviewing evaluation technique with 6 to 8 year old children. The study examines whether children are able to answer the post-task questions referring to the various interaction stages (planning, translation and assessment), and whether the technique does not lead to adverse effects such as finding a different set of interaction difficulties. The results show that children overall are fairly good at answering the questions, but have most trouble answering the planning question. Furthermore, the negative side-effects of applying the technique on the outcome of the usability test are minor. Overall, we advise practitioners to apply such a technique to uncover extra data about possible causes for interaction difficulties and to optimize the effort by only asking detailed questions about those parts of the design that need extra attention.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation (e.g., HCI)]:
User Interfaces - *evaluation/methodology*.

General Terms

Design, Human Factors.

Keywords

Usability evaluation methods, children.

1. INTRODUCTION

Various studies have elaborated on the problems of using verbalization techniques in usability tests with children. Some of these problems are summarized below. Subsequently, we describe how we dealt with these problems in a usability study on a toy product for children aged 6-8 years. This paper presents our experiences with using a structured way of

interviewing children in between given task scenarios.

1.1 Problems of Verbalization by Children

Several studies have shown that thinking aloud does not work well with children in usability tests [9, 12, 17]. Markopoulos and Bekker [15] suggest that some of the reasons for this may be children's degrees of extroversion skills and their not being accustomed to speaking up spontaneously (especially to adults). Furthermore, like adults [5], children may stop thinking aloud once an interaction gets too demanding. The problem is that thinking aloud requires verbalizations concurrently with task performance. Children in the age of 6-8 years are still in the process of developing the ability to perform tasks in parallel, as a consequence of their developing skills for abstract and logical thinking [8, 15]. In practice, this leads to children giving priority to one of the tasks (either verbalizations/co-operation, or task performance). Consequently, the number of verbalizations during task performance may be very low. Solving such problems by prompting children during task performance may lead to them mentioning problems to please the experimenter [6, 16].

Techniques like co-discovery [7, 14] also have their problems when applied to usability tests with children; Als et al. [2] report that children of 13 and 14 years of age participating in co-discovery sessions tend to discuss topics unrelated to the product under evaluation. In another study [17] with seven year olds it was found that co-discovery did not provide many verbal comments. Again, this could be due to asking the child to perform parallel tasks (i.e., co-operating and performing the task scenario).

Retrospective techniques do not rely on children's spontaneous verbalizations and on verbalizations concurrent with task performance; instead, questions are asked after task performance. Van Kesteren et al. [17] found that with retrospective techniques children proved to be able to verbalize their experiences. However, retrospective techniques do have other problems. For example, children aged 6-8 years find it difficult to keep concentrated for a longer period of time [1, 4, 17]. The problem with retrospective techniques is that sessions may last too long; sometimes twice as long as they would without the interviewing.

In an attempt to avoid the problems mentioned above, we explored the use of a structured retrospective interviewing technique.

1.2 Structured Interviewing of Children

The purpose of conducting user tests is to find usability problems in the interaction with the product. Cockton et al. [13] distinguish between four components for describing usability problems: a problem's cause, a possible breakdown or difficulty in the user's interaction, and an outcome, all of which occurs in a context. While user observations can uncover interaction difficulties, it is sometimes difficult to determine their cause(s). Verbal information provided by users can provide extra information about possible causes for the problems.

In the structured interviewing approach that we used in our study children were interviewed directly following each single performance of a given task scenario. We assumed that this would help children in not mixing up experiences from different tasks. Also, we expected that this would reduce the chance of forgetting experiences and would reduce the need of having to spend time on refreshing the child's memory by using video or other probes. Furthermore, the technique structures the interview through the use of standardized, short questions based on Andre et al.'s User Action Framework (UAF) [3]. All questions in the interview relate to three out of the four stages of the interaction process as described in the UAF [3] (Figure 1); these are the stages that directly relate to mental activities, namely: planning, translation and assessment (the fourth stage refers to physical actions). The questions are related to the different stages of the interaction cycle in order to facilitate categorising and interpreting causes for usability problems.

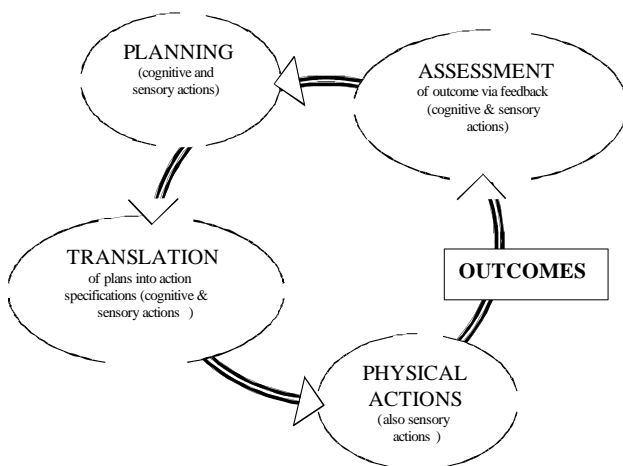


Figure 1. Illustration of the root of the User Action Framework (adapted from [10]); the stages in the interaction cycle are the starting point for categorising and interpreting causes for usability problems.

Questions were based on the following three general questions (translated from Dutch):

- 1) After I explained the task to you, what were you planning to do? (Planning)
- 2) How did you want to accomplish that? (Translation)
- 3) What happened when you did what you planned? (Assessment)

The first question refers to the planning stage of an interaction cycle. It verifies whether a child is trying to achieve an outcome

that conforms to the task goal (or whether it was just exploring the product, or trying to do something else). When a child is not trying to achieve the task's goal and the experimenter is not aware of this, the observational data can be misinterpreted. The answers to this question can be useful to understand whether the difficulties in the planning stage of the interaction are caused by the designed interaction or the planning strategies of the child. This can lead to different redesign strategies.

The second question refers to the translation stage and examines the child's planned sequence of actions needed for achieving the task goal. Pure observations do not always suffice to correctly interpret behaviour, for example, in cases where actions are considered but not performed. The answers on the translation questions can provide information about difficulties in interpreting the hints that the product provides and which hints in particular are not understood by the child.

Finally, the third question refers to the assessment stage. Questions like these are important to determine how children assess their actions and the resulting product feedback. The answers can be useful to determine whether a child has difficulties in perceiving the feedback or in interpreting the feedback. Especially when a child did not succeed in a task, this question is valuable to understand which part of the designed interaction caused the task failure.

For each specific task, the general questions as specified above were tailored so that they referred to the specific task that was given to the child. For example, in one case question one was re-formulated as: "When I asked you to turn on the toy, what were you planning to do?"

1.3 Goal of the Present Study

The goal of the study was to gain insight into the qualities and possible problems of using the structured interview approach as described above. We wanted to know whether children would be able to answer the questions in a useful and appropriate way: for example, can they tell something about their mental 'translation' process, when asked for? Is this information useful in the context of the usability test, for example, in terms of determining possible causes for the problems? Also, we wanted to explore whether applying the technique would lead to any adverse effects in comparison to task performance that is not alternated with asking questions: for example, would this lead to different sets of difficulties in the interaction? Would the children gradually gain a better understanding of the product, because of all the questions that are asked about their performance? Would they be able to perform more tasks successfully?

In the next section the setup of the study is described along with the data analysis procedure used.

2. CASE STUDY: JAMMIN' DRAW™

A toy named Jammin' Draw™, a musical drawing toy from Fisher Price (Figure 2), was used as the object of evaluation. Through sensor technology embedded in the drawing area, this toy can provide music from various musical instruments and background music styles once a child starts drawing on a paper that is on top of it. In addition, it records the music and is capable of playing it again. The toy was new to Dutch children because it was not sold in the Netherlands; none of the children had any experience with it.

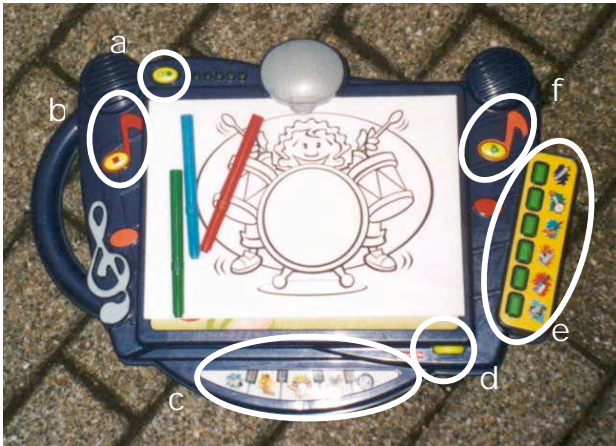


Figure 2. Jammin' Draw from Fisher Price. Explanation of buttons: a. on/off; b. record music (automatically erases previous recording); c. change background music style; d. change volume; e. change musical instrument; f. play recorded music.

2.1 Test Procedure

Two types of sessions were conducted: one in which task performance was alternated with structured interviewing, and one in which this was not the case. Each child took part in only one of the session types. In all sessions, children were asked to perform a number of tasks that together covered the complete functionality of the toy: on/off, volume control, change of instruments, change of background music style, playback and erase. While using the functions, the children coloured a picture. For each function, one single task was formulated, except for the on/off function and the volume control function, which were combined into one task description. After having performed all tasks, the children were asked to explain the various functions of the product. This was done to be able to compare the levels of product understanding across session types in search for indications of the influence that asking questions might have on a child's product understanding. A test protocol was formulated to assist the experimenter leading the sessions. The protocol contained instructions for prompting and for taking decisions about repeating a task as well as for timing the various steps. Each session took about 30 minutes. Tasks and questions are shown in Table 1. Note that both after Task 4 and Task 5 an extra assessment question was added to verify how children had interpreted the feedback.

2.2 Participants and Setting

Twenty children participated in the study. For half of them tasks were alternated with structured interviewing (see Table 2). Each child participated in only one session. The children were recruited from two school classes (class 3 and 4 of the Dutch system), which means that they were aged 6-8.

The children evaluated the toy in a small room in the school. They were randomly assigned to session types in the following

Table 2. Characteristics of test participants.

| Characteristics of children participants | Sessions with structured questioning | Sessions without structured questioning |
|--|--------------------------------------|---|
| No. of participants | 10 | 10 |
| Age in months Mean (std. dev.) | 85 (6) | 85 (7) |
| No. of boys/girls | 3/7 | 5/5 |

Table 1. Summary of test procedure (translated from Dutch).

| Test procedure |
|---|
| Child plays with toy without instructions for 5 minutes |
| <i>Task 1.</i> |
| Turn on the toy and reduce the volume |
| <i>Questions:</i> |
| Planning (P): When I asked you to turn the toy on and to reduce the volume, what were you planning to do? |
| Translation (T): How did you want to do that? |
| Assessment (A): What happened when you did that? |
| <i>Task 2.</i> |
| Colouring picture and change instruments (5 minutes) |
| <i>Questions:</i> |
| P: When I asked you to start colouring, what were you planning to do? |
| T [*]): What did you think would happen when you would start colouring? |
| A: What happened when you started colouring? |
| P: When I asked you to choose another musical instrument, what were you planning to do? |
| T: How did you want to do that? |
| A: What happened when you did that? |
| <i>Task 3.</i> |
| Choose other background music |
| <i>Questions:</i> |
| P: When I asked you to choose another background music, what were you planning to do? |
| T: How did you want to do that? |
| A: What happened when you did that? |
| <i>Task 4.</i> |
| Play back the music recorded |
| <i>Questions:</i> |
| P: When I asked you to play back the music made, what were you planning to do? |
| T: How did you want to do that? |
| A1: What happened when you did that? |
| A2: How did you know that the music was played back? |
| <i>Task 5.</i> |
| Erase the music |
| <i>Questions:</i> |
| P: When I asked you to erase the music, what were you planning to do? |
| T: How did you want to do that? |
| A1: What happened when you did that? |
| A2: Did you think that the music was erased? |
| Questions to check <i>understanding of the different functions.</i> |
| For each function: Can you tell me what this is? |
| [*]) This question was eliminated from the analysis because in hindsight it was considered that the question was not referring to the intended interaction stage (translation). |

way: the experimenter made up which type of session was next and subsequently the teacher chose the child (without knowing what the experimenter had decided). The children's behaviour and performance were videotaped. A video camera was positioned in such a way that both Jammin' Draw™ and the actions of the child were recorded (Figure 3).



Figure 3. Settings for the evaluation sessions.

2.3 Session Analysis Procedure

Recorded sessions were analysed to identify difficulties in interacting with the toy. For identifying these difficulties an adapted version of the DEVAN checklist was used [18] (see Table 3 for a part of the checklist). The checklist defines types of events that may signal difficult situations in interactions. In other words it helps the researcher recognize those events that indicate when users have difficulties in their interactions, by defining such events.

The checklist distinguishes between events based on a participant's actions on the product and expressive signals, like verbal utterances. A single difficult situation may be indicated by multiple types of events (i.e., multiple checklist items). For example, when a child presses a wrong button (ACT), indicates in words that it recognizes the action as an erroneous action (REC) and then corrects it (CORR), three signal events refer to a single difficult situation (ACT, REC and CORR respectively, see Table 3).

Two adjustments to the original DEVAN checklist were made in order to tailor it for use in the study. In the original checklist, the indication 'wrong action' (ACT) represents an observed action on the product. However, in the present study it was also used in cases where children brought up a wrong choice of action in their answering of questions (expressive signal). Also, in some cases children explained actions of which the interviewer knew they were wrong while the children did not realize they were wrong. In such cases the code ACT was used as well. Finally, another adaptation was that children had two chances to correctly do a task before a wrong action was coded as such (ACT).

After having combined those signal events that referred to a single difficult situation, the difficult situations were listed for each separate session. The signals themselves were marked as coming either from a task performance part or an interviewing part of a session.

Finally, based on the session-specific lists of difficult situations, the third author of the present paper described the difficulties in

Table 3. Part of the checklist of difficulty signals, as an example of how signal events are defined.

| Short description | Definition |
|---|---|
| ACT: Wrong action | (based on action on product or on user expression in structured interview) an action does not belong in the correct sequence of actions an action is omitted from the sequence an action within a sequence is replaced by another action actions within the sequence are performed in reversed order (based on user expression in structured interview) user mentions to have performed a wrong action user mentions action of which the analyst knows it is wrong, but the user does not realise this |
| CORR: Corrective action | (based on action on product) an action is corrected with a subsequent action (or sequence of actions) an action is undone |
| REC: Mis-understanding or recognition of error | (based on user's (verbal) expression) user indicates: to recognise a preceding error to understand something previously not understood |

the interaction with the product, thus constructing one master list of difficulties. To determine the reliability of the coding scheme in terms of any-two agreement (agreements/ total number of difficulties found), four sessions out of 20 were coded by another evaluator (Table 4). The average any-two agreement (50%) was similar to other studies described in this area [11]. Furthermore, the discussions between the two evaluators led to a better description of the coding scheme's rules.

Table 4. Number of agreements, number of unique observations and number of disagreements between two evaluators.

| Eval A * Eval B | Any-two | Agreements | Unique by A or B |
|--------------------|---------|------------|------------------|
| S1, SI | 47% | 9 | 10 |
| S2, SI | 53% | 8 | 7 |
| S3, WSI | 38% | 5 | 8 |
| S4, WSI | 60% | 9 | 6 |

3. ANALYSIS AND RESULTS

3.1 Can Children Answer the Questions?

To find out to what extent the children participants were able to answer the structured interviewing questions appropriately, we determined whether the answers referred to the intended interaction stage. For example, in a planning answer children should mention what their goal was; in a translation answer children should mention a range of specific actions, for example, to press a button and to start colouring again; in an

assessment answer the child should mention the feedback given by the toy, for example, a sound. The results are presented in Table 5.

A Chi-square analysis showed a significant association between questions and answers ($df = 6, \alpha \leq 0.001$). A post-hoc analysis of the standard residuals of the (grey) planning, translation and assessment cells showed that the obtained frequencies in these cells were all significantly above the expected frequencies ($\alpha = .004$, with z scores of 6.7, 5.4, and 7.2, respectively). The answers related significantly more frequently to the correct (i.e., intended) than to the incorrect (not intended or undefined) interaction stage according.

Table 5. Overview of the answers to post-task questions indicating whether the answers refer to the intended interaction stages.

| Questions related to interaction phase... (number of questions per child) | Answers in terms of... | | | | Total |
|--|------------------------|-------------|------------|-----------|-------|
| | Planning | Translation | Assessment | Undefined | |
| Planning (6) | 34 | 19 | 1 | 6 | 60 |
| Translation (5) | 2 | 36 | 0 | 12 | 50 |
| Assessment (8) | 0 | 2 | 69 | 9 | 80 |
| Total (19) | 36 | 57 | 70 | 27 | 190 |

The ‘planning’ questions (What were you planning to do?) often elicited answers referring to the translation phase of the interaction process (see Table 5). Instead of answering: “To turn on the toy” or “To erase the music made”, some children, for example, said: “I was planning to press the buttons”. For most children this occurred at least once during each session. However, in many cases their answers did reflect the planning interaction phase. Children also answered with: “To do it” on a planning question, which is not that strange considering the fact that one of the questions for example was: “What were you planning to do when I asked you to erase the music”. Answering with “To do it” is an answer referring to the right action stage, although it is not a very informative answer.

In response to the ‘translation’ questions (How did you want to do that?), children often provided “I don’t know” answers. Especially in cases where a child answered the planning question with a ‘translation answer’, the subsequent question which referred to the translation phase was difficult to answer. In response to translation questions children also answered: “Just press buttons”, without explaining which buttons. This can indicate that they had a goal, but their translation was just to look and try until they succeeded in the plan. Therefore, the answer refers to the action stage it aims at (i.e., translation).

‘Assessment’ questions (What happened when you did what you planned?) mostly elicited answers about the assessment phase. For example: “You hear bleep bleep”, “Then the music came”, or “Now, these lamps started and then you often hear a sound of these or this one, of the instruments”. In their answers the children often mentioned only one of the elements they could have assessed, even though after most actions, more than one thing happened. For example, when pressing an instrument button, a short sound was heard and the newly chosen instrument could be heard when the child started colouring again. It was unclear why children often mentioned only one

type of feedback. It is possible that the children actually did assess all feedback but only mentioned one in their answers, for example, because the question made them come up with short answers, or because the children were not used to give a series of answers. Another explanation would be that the product did not provide enough feedback and that the children indeed assessed only one type of feedback.

A similar result was found in the question part after the first task. The questions after the first task asked about two subtasks in the same question (When I asked you to turn the toy on and to reduce the volume, what were you planning to do?). Most children answered about how they interacted during only one of these subtasks. Only one child answered both sub-questions, namely: “You can have it make louder or less loud sounds, like this and here ‘on and off’”.

On only 27 of the total of 190 questions children were not able to answer or were very unclear in their answer. They mostly answered “I do not know”, or paused for a long time. All children had difficulties with one or two of the questions although they were able to answer the other questions. The translation questions of the first two tasks led to most undefined answers, probably because the children answered the preceding question in terms of translation.

The children often used the toy to explain what they did or thought. They pointed out the buttons or repeated the action to stress their answers. They answered, for example: To press here”, or: “Just press this one.” Occasionally the child describes the button, for example: “These green ones” (about the instrument buttons), or: “The one with the square” (about the new song button). The presence of the object that is studied thus seems to be important for the children’s ability to answer the questions.

3.2 What Extra Information do the Answers Provide?

Figure 4 shows the difficulties identified in the sessions with and without the structured interviews. The numbers in the cells signify participant identification numbers (numbers run until 22 instead of 20 because sessions of participants 13 and 19 were cancelled due to technical problems with recording them). In the sessions with the structured interviews 22 interaction difficulties were found, whereas the sessions without interviews revealed 18 interaction difficulties.

In the following we discuss the answers given during the structured interview sessions to analyse the extra information revealed by the technique, for example, in terms of possible causes of the problems. We provide an example of each task.

During the first task, interaction difficulty 2 (finding the on/off button) was found in four of the structured interview sessions. In three of these sessions an indication of the difficulty was found in the answers to the questions. The first participant tried to press the new song button to turn the toy on. The answers on the questions were: “Then I thought, is it this or this (about the play back and new song button)”, “Actually I wanted to do it here” (points at the new song button), and: “But I did this (presses the on/off button) and then the music appeared”. The answers confirm that the child had trouble finding the on/off button and shows that the child understood that the new song button was the wrong one. Participant 7 pressed several buttons before she found the on/off button. Her answer on the planning question revealed that she did not have a difficulty in finding the button, but that it was part of her strategy. She answered: “I was planning to look whether this one works (new song button),

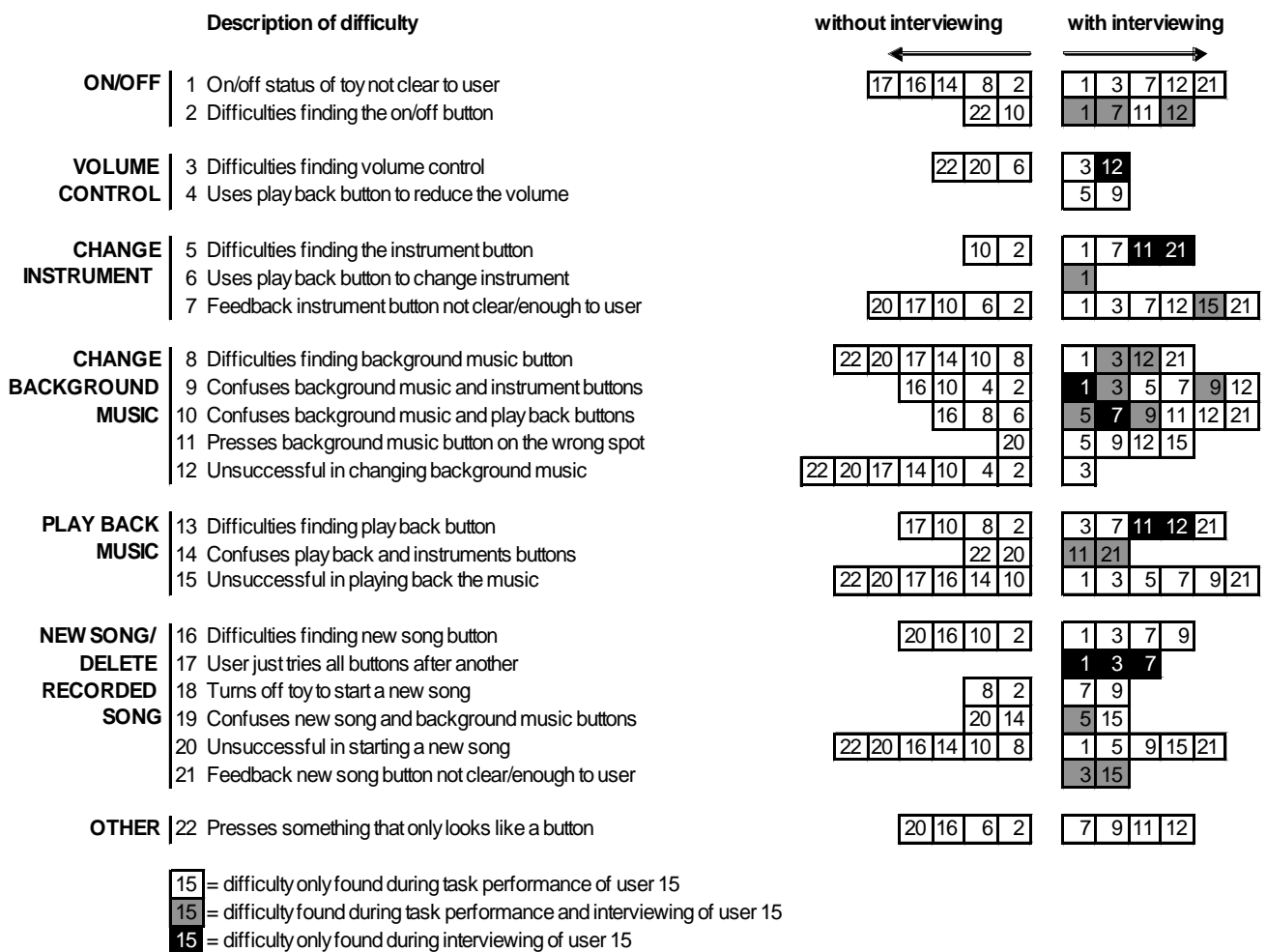


Figure 4. Difficulties identified in each of the sessions.

and that one (play back button) and then I went here (on/off button)”. The twelfth participant indicated in his answers that he just did not know how to start. In this example the extra information provided by the technique is that although the children had difficulties in finding the on/off button, the feedback provided by the button is sufficient and probably no adjustments need to be made in the design of this button.

The fifth interaction difficulty (finding the instrument button) was found in two sessions during the task part and in two other sessions only during the questions part. Participant 11 was pressing the right buttons to perform the task given, however on the translation question the participant answered: “Here are all kinds of little music (about the instruments buttons), but all the time the same one”. She thus appeared to have difficulties in finding the right buttons because there was no adequate feedback to confirm that she was right. This was not revealed during the observation of the task. Participant 21 also pressed the right buttons during the task part, but explained during the questions that he was not able to find the right buttons at first. Thus, the difficulty with these buttons is that they do not adequately show what their function is and do not provide adequate feedback about their activation. Only observing the task performance in session 11 and 21 could not have resulted in finding this difficulty with the instrument buttons.

In the background music task many difficulties were revealed during the task part and the question part. During the task part the children pressed different buttons to try to change the background music. The answers to the questions revealed that some of the children confused not only the buttons, but also the concepts of instruments and background music. The instruments could only be heard when after pressing that button the picture field was pressed, while the background music is constantly played, even when not pressing on the picture field. This difference was difficult for the children. They answered, for example, (participant 1): “I was planning to do this one (about the instrument buttons)”, then I wanted to press that one (instrument button) and include this one (play back button)” and: “yes a little music came I thought”. This child did not only confuse the buttons but also interpreted the feedback incorrectly. A similar example was found in the answers of participant 9: “Another background music”, “With this one (instrument buttons)”, and: “Then a music came”. Remarkable in the background music task is that none of the children evaluated during the questions that they pressed the background music buttons on the wrong spot.

The task of playing back the music resulted in difficulties of finding the right button and confusing the play back button with the instruments buttons. Participant 11 and 12 both explain in their answers that they had to search for the right button, although during the observations no difficulties were found.

Participant 11 however used the wrong button for the task, which was revealed both during the task and the question part. The answers did not provide extra information here. The same holds for the answers of participant 21.

During the last task a difficulty was found only after analysing the answers of the children (difficulty 17). The answers of participants 1, 3 and 7, leading to describing this difficulty, were: "Then I wanted to press this one (volume control), then I wanted to press this one (on/off button) and then I pressed this one (instrument button)", "At first this one (instrument button), then by this one (background music button) and then by this one (unclear which button), just look which one was right", and "Try something here (play back button) and then this one (instrument button) and then I had that one (on/off button)". These answers indicate that they were trying to find the right button, but also show that two of the children thought they were successful although they were not. Especially, the children's assessment of task success point towards difficulties that can not be revealed by only observing the behaviour during task performance.

The answers to the questions can provide extra information, for example, when a child experiences difficulties in finding the right buttons, but does succeed in his or her task. Only observations would not have resulted in finding these difficulties about the hints that a button gives about its function. Especially the sequence of answers is informative. When a child uses wrong buttons during a task, the answers do not provide so much extra information, it just repeats the observation.

3.3 Does the Interview Influence the Interaction with the Product?

To determine whether applying the structured interviewing technique leads to negative side-effects we compared the outcomes of the sessions with structured interviewing to those without structured interviewing.

3.3.1 Influence on Difficulties Set

Of the four difficulties identified in the sessions with structured interviewing, only one (difficulty 17) was identified in the interviews only (Figure 4). The fact that this difficulty was identified in the interviews only, is due to the fact that, strictly, by just observing users one can not be sure that they are 'just pressing all buttons one after another' (after all, there may be some plan behind it). Interviews provided clarification on this. Furthermore, the answers to the questions mainly provided extra information on what they were trying to do (e.g., "no I didn't know it so well, how to reduce the volume" (child 12, difficulty 3); "...try this one, then this one, then that one, and just look which one is right" (child 3, difficulty 17); or how they interpreted feedback (after the question: "did you think the music was now gone?" - "I don't know" (child 3, difficulty 21).

3.3.2 Influence on Product Understanding

After having performed all tasks, children were asked to explain each of the product's functions. If the children that participated in the sessions with structured interviewing would have better explanations, this indicates that their understanding of the product is better. The answers of the participants were categorized based on the quality of the answer. The categories were: 1) correct explanation, 2) unclear explanation (unclear whether the function was properly understood) and 3) wrong explanation and "I don't know" answer. Table 6 indicates that the children in the sessions with structured interviewing had a

slightly better understanding of the product functions than those participating in the sessions without structured interviewing. However, the difference was not significant according to a Chi-square test ($df = 2, \alpha \geq 0.05$).

Table 6. Explanation of functions at the end of the session.

| Explanation category | Nr. of answers per category | |
|------------------------------------|-------------------------------|----------------------------|
| | Without structured interviews | With structured interviews |
| Correct explanation | 24 | 31 |
| Unclear explanation | 9 | 10 |
| "Don't know" and wrong explanation | 37 | 29 |
| Total | 70 | 70 |

3.3.3 Influence on Task Success

Table 7 shows the number of tasks that were performed successfully without any difficulties, successfully with difficulties, and unsuccessfully. It does so, both for the sessions with and without the structured interviews in between tasks. Table 7 shows that children with structured interviews were slightly more successful in their task performance. However, this difference was not significant according to a Chi-square analysis ($df = 2, \alpha \geq 0.05$).

Table 7. Task performance with and without structured interviewing for five tasks.

| Task performance | Without structured interviewing | With structured interviewing |
|-----------------------------------|---------------------------------|------------------------------|
| Successful, without difficulties | 7 | 6 |
| Successful, but with difficulties | 22 | 26 |
| Unsuccessful | 21 | 18 |

4. DISCUSSION AND CONCLUSION

The results show that children were able to answer most of the structured interview questions. However, in some cases, planning questions lead to answers not related to the planning, but rather to the translation stage of actions. In some cases questions referred to different subtasks or to a series of feedback occurrences. In such cases the children often did not give a series of answers, but gave only one answer. In the study, we kept to a strict protocol to be able to evaluate the technique. However, children of this age group might be better able to answer a series of shorter questions. The experimenter can then prompt the children to explain their interaction steps one by one.

No significant effects of structured interviewing were found on understanding of functions, nor on task success. In terms of number of difficulties identified, the results show that in the sessions with structured interviewing more difficulties were identified than in the sessions without.

Two possible benefits of structured interviewing can be finding more indications of difficulties, which increases the chance that

a difficulty is found, and getting better information about possible causes of difficulties because the explicit comments by the children require less interpretation than observing (non-verbal) actions during task performance. For example, in the case of one of the difficulties found in structured interviewing sessions (difficulty 17, which occurred with three users) the structured interview technique revealed user intentions (“just trying all buttons one after another”) that remained hidden without having the interviews (i.e., the DEVAN technique used to analyse the observations only allows for identifying such intentions if there is an explicit mentioning of them).

The structured interviewing technique contributed to uncovering 13 out of the 21 difficulties. However, because of the fact that we applied a fairly clean question protocol the verbal comments did not always contribute to a better understanding of the difficulty. The structured questions provide a first step in categorising and interpreting questions in terms of the interaction cycle phase, however subsequently more detailed questions can uncover possible causes of the difficulties, for example, whether the icon on the button is unclear, or whether it is unclear what parts of the product are buttons. Some further prompting questions could have been asked to elicit the reason for the children’s difficulties. Future research should examine how the question protocol can be adapted to optimise the quality of the children’s comments.

4.1 Practical Advice

What would we advise practitioners when they are planning to evaluate products with young children given our experiences in this study? The protocol that was used for the structured interviewing was partly determined by the fact that we were examining not only the product but also the technique itself. The structured interviewing technique can be a useful technique to elicit extra verbal information from children that will allow an evaluator to interpret possible causes for an observed difficulty. However, to optimize the evaluation effort we would suggest that questions are asked only after tasks for which the observer feels there is a need for additional information to better understand the user’s actions or difficulties. Furthermore, children will provide more useful information when a less ‘clean’ protocol is followed and extra prompting questions are asked to elicit more details about difficulties that they experienced.

5. ACKNOWLEDGMENTS

The study would not have been possible without the help of the children. We acknowledge their input and also the support of the teachers, who offered a quiet place in school for testing the product. We also want to thank Wolmet Barendregt for coding the four sessions to determine the agreement between different evaluators.

6. REFERENCES

[1] Acuff, D. S., and Reiher, R. H. *What Kids Buy and Why: The Psychology of Marketing to Kids*. The Free Press, New York, 1997.

[2] Als, B. S., Jensen, J. J., and Skov, M. B. Exploring verbalization and collaboration of constructive interaction with children. In *Proceedings of the IFIP 9th international conference on Human-computer interaction (INTERACT 2005)*. Springer, Berlin, 2005, 443-456.

[3] Andre, T. S., Hartson, H. R., Belz, S. M., and McCreary, F. A. The user action framework: A reliable foundation for usability engineering support tools. *International Journal of Human-Computer Studies*, 54 (2001), 107-136.

[4] Berk, L. E. *Child Development*. Allyn and Bacon, MA, 1997.

[5] Conrad, F., Blair, J., and Tracy, E. Verbal reports are data! A theoretical approach to cognitive interviews. In *Proceedings of the 1999 federal committee on Statistical methodology research conference*. (Arlington, VA, 1999). 1999, 11-20.

[6] Donker, A., and Reitsma, P. Usability testing with young children. In *Proceedings of the 2004 conference on Interaction design and children*. ACM Press, New York, NY, 2004, 43-48.

[7] Dumas, J. R., and Redish, J. C. *A Practical Guide to Usability Testing*. Ablex Publishing Corporation, Norwood, NJ, 1993.

[8] Greig, A., and Taylor, J. *Doing Research with Children*. Sage, London, 1999.

[9] Hanna, L., Risdien, K., and Alexander, K. Guidelines for usability testing with children. *Interactions*, 4, 5 (1997), 9-14.

[10] Hartson, H. R. Cognitive, physical, sensory, and functional affordances in interaction design. *Behaviour & Information Technology*, 22, 5 (2003), 315-338.

[11] Hertzum, M., and Jacobsen, N. E. The evaluator effect: A chilling fact about usability evaluation methods. *International Journal of Human-Computer Interaction*, 15, 1 (2003), 183-204.

[12] Höysniemi, J., Hämäläinen, P., and Turkki, L. Using peer tutoring in evaluating usability of physically interactive computer game with children. *Interacting with Computers*, 15, 2, (2003), 203-225.

[13] Lavery, D., Cockton, G., and Atkinson, M. P. (1997). Comparison of evaluation methods using structured usability problem reports. *Behaviour & Information Technology*, 16, 4/5 (1997), 246-266.

[14] Kemp, J., and van Gelderen, T. Co-Discovery exploration: An informal method for the iterative design of consumer products. In P.W. Jordan, B. Thomas, B. Weerdmeester, and I. McClelland (eds.), *Usability evaluation in industry*. Taylor & Francis, London, 1996, 139-147.

[15] Markopoulos, P., and Bekker, M. M. On the assessment of usability testing methods for children. *Interacting with Computers*, 15, 2 (2003), 227-243.

[16] Nisbett, R., and Wilson, T. Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84 (1977), 231-259.

[17] Van Kesteren, I. E. H., Bekker, M. M., Vermeeren, A. P. O. S., and Lloyd, P. A. Assessing usability evaluation methods on their effectiveness to elicit verbal comments from children subjects. In *Proceeding of the 2003 conference on Interaction design and children*. ACM Press, New York, NY, 2003, 41-49.

[18] Vermeeren, A. P. O. S., den Bouwmeester, K., Aasman, J., and de Ridder, H. DEVAN: A tool for detailed video analysis of user test data. *Behaviour & Information Technology*, 21, 6 (2002), 403-23.

