

# Empowering Teenagers to Perform a Heuristic Evaluation of a Game

Obelema Akobo Wodike  
ChiCI Group  
University of Central Lancashire  
Preston, UK  
oakobo@uclan.ac.uk

Gavin Sim  
ChiCI Group  
University of Central Lancashire  
Preston, UK  
grsim@uclan.ac.uk

Matthew Horton  
ChiCI Group  
University of Central Lancashire  
Preston, UK  
mplhorton@uclan.ac.uk

**Unlike user studies, inspection based methods are not widely researched in the area of Child Computer Interaction. This paper reports the findings of a study to empower teenagers to facilitate a heuristic evaluation with their peers acting as the expert evaluators. In total 20 teenagers participated in the study, with four of the teenagers acting as facilitators and the remainder as evaluators. The results showed that teenagers struggled to act in the role as facilitator, struggling to explain the heuristic evaluation process and keep the evaluators on track. The evaluators found very few problems and became distracted from the evaluation opting to play on other features of the device rather than the game itself. Further research will be performed to modify the process in an attempt to eliminate these issues in order to improve the method for teenagers.**

*Heuristic Evaluation, Children, Teenagers, Games, Usability.*

## 1. INTRODUCTION

Within the HCI community there has been considerable research conducted on evaluation methods (Polson et al., 1992, Jeffries et al., 1991) and these methods have been adapted for use with children by the Child Computer Interaction (CCI) community. Once adapted research has shown the effectiveness of many traditional evaluation methods (Read, 2008, Baauw and Markopoulos, 2004, Zaman and Abeele, 2010, Sim et al., 2005) for use with children. The majority of evaluation methods used within CCI are user based and inspection methods have largely been overlooked. The SEEM method was developed (Bauuw et al., 2006), yet this used adults as expert evaluators who examine a game answering a checklist with a series of questions. A limited number of studies have attempted to use children to perform a heuristic evaluation (MacFarlane and Pasiali, 2005).

One of the requirements in order to perform an effective heuristic evaluation is the need to use double experts as the evaluators (Nielsen, 1992), experts in both usability and the domain being investigated. However, within some contexts being a double expert might not be sufficient. For example there would be a need to understand children, games and usability if the product was a children's game, thus a triple expert might be necessary. Therefore using children as experts to evaluate products or technology to inform the

design is desirable because according to Zaman (2010), adults are unable to judge whether a game for children will be fun, challenging and user friendly, they have lost the feeling of being children.

Within the context of participatory design it has been argued that children are experts in the way they interact with their world and capturing this expertise is believed to be key to designing meaningful artifacts for children (Iversen and Brodersen, 2007). Children have performed different roles in the design and evaluation of technology for example stakeholders (Iversen and Brodersen, 2007), design partners (Druin et al., 1999), informants (Scaife and Rogers, 1999) or testers (Gilutz and Nielsen, 2002). Given an assumption that children can be domain experts in the use of technology it is conjectured that, given appropriate training, they might be able to effectively perform a heuristic evaluation. Research conducted with 10-11 year olds performing a heuristic evaluation identified a number of issues including their ability to understand the heuristic set and severity scales (Salian et al., 2013). These issues might be eliminated or reduced if the study was performed using older children. There is also uncertainty of the affect the training and facilitation has on the children performance.

This research aims to explore whether teenagers can be empowered to facilitate a heuristic evaluation with their peers and identify the issues they have with the evaluation process.

## 2. METHOD

For this study the aim was to observe the heuristic evaluation process in action without influencing the evaluators. Direct observation was used to capture any issues either the facilitator or the evaluator experienced.

### 2.1 Participants

In total 20 teenage boys, aged between 12 and 13, from a secondary school within the UK participated in the study evaluators. Four of the teenagers were randomly selected to act as the facilitator and the remaining teenagers were the evaluators. There were 4 groups in total consisting of 4 evaluators and 1 facilitator.

Two researchers acted as observers for the entire process. Both researchers are experienced with the heuristic evaluation process and have experience of carrying out evaluation methods with children and teenagers.

### 2.2 Apparatus

The teenage evaluators were provided with three documents consisting of the game heuristic set (Pinelle et al., 2008), severity scales and a form to record any usability problems they encountered, see Figure 1.

Ninjabo log	<input type="checkbox"/> Reading Game Instructions and story	<input type="checkbox"/> Scanning for problems	8
	<input type="checkbox"/> Navigating within the game UI	<input type="checkbox"/> Trying to force errors	
	<input checked="" type="checkbox"/> Actual game play	<input checked="" type="checkbox"/> Following users task	
	<input type="checkbox"/> Finishing the game play		

Figure 1: Evaluators form for recoding individual problems

On the form they had to record the problem, the heuristic it violated, state where in the game the problem occurred, how they found it and finally attach a severity. In the example above the evaluator had attached the heuristic number to the severity column.

For the aggregation stage of the heuristic evaluation process they were provided with a form to enable the teenager to aggregate their problems into a single list of usability problems. They were also provided with an iPad with the application preinstalled.

Following initial training the teenage facilitators were given scripts to aid them explaining the heuristic evaluation process to their peers. This consisted of an explanation of the heuristic set, severity ratings, the evaluation forms and the merging process.

The researchers had two separate forms to capture issues in the individual evaluation and a separate

form to document problems with the merging process.

### 2.3 Application

An iPad game was selected that the authors felt might appeal to the target audience and upon initial inspection the authors identified a number of potential usability problems. Lego® Ninjabo Rebooted see Figure 2. The game was a free running game in which the character continually runs, having to avoid objects, collect bricks and make their way to the top of the tower.



Figure 1: Ninjabo Rebooted Screen Shots

### 2.4 Heuristic Set and Severity Scales

There are a number of different heuristic sets available for evaluating games (Federoff, 2002, Desurvire et al., 2004). The decision was made to use the heuristic set by Pinelle et al. (2008) as they were deemed appropriate and with only 10 heuristics it would not overwhelm the evaluators or facilitators compared to some of the heuristics sets with greater numbers (Desurvire et al., 2004).

The decision was made to use Nielsen's severity scale (Nielsen, 1992) as there are no domain specific severity scales for games. Therefore problems were rated using the five point scale from 0 to 4, where 0 means "I don't agree that this is a usability problem at all" and 4 means "Usability catastrophe: imperative to fix this before product can be released".

### 2.5 Procedure

The teenagers came into university as part of a Mess Day (Horton et al., 2012) and were randomly allocated into groups. Each of the groups was then trained in a research method and in this case the 4 teenagers were trained in how to perform a heuristic evaluation. The two researchers (who were then carrying out the observation) spent 30

minutes explaining how to perform a heuristic evaluation whilst the teenagers acted as evaluators. Once the teenagers had completed the various stages the process was explained and the teenagers had opportunities to ask any questions. At the end the teenagers were given notes on the method and scripts to assist them in explaining the process to their peers.

Following the training session each facilitator then instructed a group of 4 or 5 teenagers in how to perform the evaluation. Each facilitator instructed their peers and if necessary one of the observers intervened to help the facilitator if they could not remember certain aspects of the process. Each of the 4 evaluation sessions were scheduled to last 30 minutes, with approximately 15 minutes set aside to play the game and the remaining time to explain the process and aggregate the data.

Before the process started the two researchers introduced themselves to the group and informed them of their role within the study. The evaluation was conducted on a large round table and the two researchers were facing the evaluators in order to capture any issues that arose.

## 2.6 Analysis

The individual evaluator's forms and the aggregated forms were analysed by the authors. In the first stage the individual forms were examined to determine whether the problems reported seemed feasible based on their experience of playing the game. Similar problems within each group were then aggregated and these were then compared to the data that was presented in the aggregated list for the group.

The observational data was broken down into two different data sets, one for the facilitator and the other for the evaluators. The data was analysed by the authors of the paper to determine common themes that emerged within the groups. These themes could then be used for further research to assist in the redesign of the evaluation method.

## 3. RESULTS

Two sets of results are presented, the results obtained by the evaluators performing the heuristic evaluation and the observational data.

### 3.1 Evaluation Results

In total 14 out of the 16 evaluators reported at least one usability problem on the evaluation sheets provided, with 17 problems being reported. Only 3 of the evaluators actually reported more than a single problem, they documented two problems each. However not all problems documented appeared to be genuine problems or easy to

interpret, for example the evaluators reported the following issues:

- No Problem
- Attack
- Menu

For the first problem the evaluator suggested it violated the heuristic 'Provide users with information on game status' and gave it a severity rating of 3, demonstrating potentially little understanding of the process. The second comment was the only information provided on the evaluation form, no severity rating or heuristic it violated was documented. For the final problem it was suggested that it breached 'Provide predictable and reasonable behaviours for computer controlled units' but no severity scale was attached.

Of the 17 problems reported only 3 of the problems were reported accurately, in that they provided all the information including heuristic it violated and severity rating. No additional information other than the problem was noted in 7 instances and there were cases with inaccurate severity ratings being provided, for example one evaluator gave it a rating of 8.

Despite these issues some genuine problems did appear to be reported, for example:

- Don't know how to kill the ninjas (reported 6 times)
- Doesn't react straight away sometimes when I click (reported 4 times)
- How to duck under red things in your way (reported 2 times)

For the merging stage where the groups aggregated their individual problems into a single list, these are shown in table 1. The total problems represented the number of individual problems reported by the evaluators within the group; merged problems is the number of problems documented on the aggregation form; added problems are new problems that only appeared on the aggregation form and missing problems are those which were on the individual forms but did not appear on the aggregated list.

**Table 1:** Number of problems reported and aggregated

Group	Total Prob.	Merged Prob.	Added Prob.	Missing Prob.
A	4	1	0	3
B	4	2	0	1
C	4	1	0	1
D	5	7	2	0

Three of the groups failed to aggregate all the problems into a single list. Whilst group D, added 2 additional problems that had not been previously documented on any of their individual forms. Upon close inspection there were also inaccuracies with the frequencies of discovery reported on the forms. For example in group C, 3 evaluators reported the fact that they did not know how to kill the ninjas but on the aggregated form the frequency only showed two. In contrast only 1 person reported minor lag, in group A, but the frequency of discovery was reported as 2.

### **3.2 Facilitator Observational Results**

There were a number of problems observed during the studies in which the facilitator struggled and had to be assisted by one of the researchers. Despite the initial training and a document being produced that the teen could read out explaining the process all the facilitators struggled to explain the process. All four groups had to have one of the researchers intervene to explain either the forms, the heuristic set, severity scales or the tasks. It was observed in two of the groups that the evaluators were not actually listening to the facilitator when they were trying to explain the process. This occurred in the merging stage of the heuristic evaluation, as the teenagers had become disengaged with the evaluation at this stage.

Other problems that were observed include:

- Telling people what to write on the forms
- Forgetting to tell the evaluators to merge the problem set
- Playing with the camera on the iPad
- Left the activity and came back a few minutes later
- Assisting the people play the game and actually playing the game instead of the evaluator

In one of the groups, as the evaluators hadn't actually documented any problems, the facilitator informed the evaluators that they should write down some of the problems such as Lag that were identified in the earlier training session. It became clear that the facilitator became disengaged with the process in three of the four groups. This may have been attributed to the fact that the evaluators had stopped performing the evaluation and were simply playing with the iPad. In two separate groups the facilitator tried to encourage the evaluators to continue the activity but they appeared to be having too much fun to listen. Therefore the facilitator at this point seemed to abandon the evaluation altogether and joined in with the evaluators.

### **3.3 Evaluators Observational Results**

The evaluators struggled to perform a heuristic evaluation for a number of reasons. The facilitator struggled to initially explain the process but even though the researchers explained the process all the groups abandoned the evaluation at different stages.

As stated above, the teenagers were not listening to the facilitator and two groups did not look at the heuristic set or the severity scales. Even when they looked at the heuristic set they did not understand them and asked for assistance. All the evaluators struggled to complete the forms even when they were initially engaged with the activity.

All the evaluators stopped playing the game after a few minutes and started to play with other features on the iPad. For example two of the groups found music and started playing this, which distracted the other group members, who then started playing with other features. The camera was also popular with the evaluators taking pictures of their peers and selfies. One evaluator managed to take over 50 pictures in the session.

Other less frequent problems were:

- Telling other evaluators what to write down
- Not recording problems when they were identified (they had ask for assistance)

In one of the groups an evaluator would find a problem and then would inform all the other evaluators who would then write this on their own forms. Clearly this would skew any results from the evaluation.

## **4. DISCUSSIONS AND RECOMMENDATIONS**

The process in which the teenagers were trained to instruct their peers to perform a heuristic evaluation was clearly not a success for a number of potential reasons.

Despite the fact that the facilitators took part in a heuristic evaluation and were trained in this, they all struggled to inform their peers of the process. Even when instructions are presented they did not use them. It is questionable whether they had sufficient training to explain the method or even if substantial training had been provided, whether they would have felt comfortable explaining the method.

The facilitator also struggled to keep the activity on track. It might have been that the evaluators did not see the facilitator as an authority figure and thus ignored instruction and just played with the devices. When the researchers intervened the evaluators would listen at this point.

The role of a facilitator might not have been much fun compared to the evaluators who were allowed to play with the iPads. There may be a need to ensure that the activities are fun for all participants in future studies.

The evaluators might have struggled to differentiate the process of evaluating a game and just playing the game. Most of the evaluators were discussing their highest scores amongst themselves and appeared to disengage from the evaluation process. Once they had become bored with the game they then started to look for other features on the iPad to play with including the camera. There is clearly a need for the evaluators to understand an evaluation process and make a distinction about being critical of a game and simply playing the game to be entertained.

It is clear that the heuristic evaluation process is difficult for teenagers to engage with and perform successfully. In order to prevent problems encountered from reoccurring in future work some recommendations have been provided thus:

- It is potentially rather unfair for the teenager to be responsible for the facilitation. Thus facilitation should not be left solely to the teen peer. The presence of an adult in the facilitation process may make the children judge the activity as serious rather than just a fun.
- The training of the facilitators and evaluation is key to the overall success of the heuristic evaluation. An thorough and detailed training session would be useful. This may involve the adults acting out the evaluation procedure before the teen facilitators but keeping them engaged during this process may be difficult. After this process they would then conduct the evaluation to ensure they understand the method.
- Both the evaluation process and training should be made fun for all participants. It is important that the role of the teen facilitator is also engaging and they do not feel that they are potentially missing out on the fun activity of playing the game.
- At intervals during the evaluation, evaluators could be reminded to record problems encountered. If feasible it may be useful to get them to pause the game every few minutes or at certain levels to record issues found.
- The tools provided to the evaluators and facilitators e.g. heuristic set, severity scale and forms should be made more suitable for the teenagers. These forms may be

moderated by a teacher before they are given to the teenagers.

- Children should be well spaced out and be made to understand that the individual session should be done individually to prevent peer bias or influence.

## 5. FUTURE WORK

Based on the recommendations future work will look at ways of empowering teenagers to be critical of software in a playful and more user friendly way. This will include the way that problems are captured as it is clear that the forms used in this study are not suitable for teenagers. Whether it is possible to empower teenagers to act as the facilitator is questionable, it might be that adults still need to be involved in order to assist, provide explanations and ensure that the evaluators do not digress from the activity.

## 6. REFERENCES

- Baaui, E. & Markopoulos, P. (2004). A comparison of think-aloud and post-task interview for usability testing with children. *IDC 2004*. ACM Press: Maryland.
- Baaui, E., Bekker, M. M. & Markopoulos, P. (2006). Assessing the applicability of the structured expert evaluation method (SEEM) for a wider age group. *Interaction Design and Children*. Tampere: ACM.
- Desurvire, H., Caplan, M. & Toth, J. A. (2004). Using heuristics to Evaluate the Playability of Games. CHI, 2004 Vienne. ACM.
- Druin, A., Bederson, B., Boltman, A., Miura, A., Knotts-Callahan, D. & Platt, M. (1999). Children as our technology design partners. In: DRUIN, A. (ed.) *The design of children's technology*. San Francisco, CA: Morgan Kaufmann.
- Federoff, M. A. (2002). *Heuristics and Usability Guidelines for the Creation and Evaluation of Fun in Video Games*. MA, Indiana University.
- Gilutz, S. & Nielsen, J. (2002). *Usability of websites for children: 70 design guidelines* [Online]. Nielsen Norman Group, Fremont. Available: <http://www.nngroup.com/reports/kids/>.
- Horton, M., Read, J. C., Mazzone, E., Sim, G. & Fitton, D. (2012). School friendly participatory research activities with children. *CHI 12 Extend Abstracts*. Austin Texas: ACM.
- Iversen, O. S. & Brodersen, C. (2007). Building a BRIDGE between children and users: a socio-cultural approach to child-computer

- interaction. *Cognition, Technology & Work*, 10, (83-93).
- Jeffries, R., Wharton, C. & Uyeda, K. M. (1991). User interface evaluation: A comparison of four techniques. CHI 91, 1991 New Orleans. ACM, 119-124.
- Macfarlane, S. & Pasiali, A. (2005). Adapting the heuristic evaluation method for use with children. Workshop on child computer interaction: methodological research, 2005 Interact 2005, Rome. 28-31.
- Nielsen, J. (1992). Finding usability problems through heuristic evaluation. Proceedings of the SIGCHI conference on Human factors in computing systems, 1992 Monterey. ACM, 373-380.
- Pinelle, D., Wong, N. & Stach, T. (2008). Heuristic Evaluation for Games: Usability Principles for Video Game Design. *CHI2008*. Florence: ACM.
- Polson, P., Lewis, C., Rieman, J. & Wharton, C. (1992). Cognitive walkthroughs: A method for theory-based evaluation of user interface. *International Journal of Man-Machine Studies*, 36, (741-773).
- Read, J. C. (2008). Validating the Fun Toolkit: an instrument for measuring children's opinion of technology. *Cognition, Technology and Work*, 10, (119-128).
- Salian, K., Sim, G. & Read, J. C. (2013). Can children perform a heuristic Evaluation? *APCHI2013* Bangalore, India.: ACM.
- Scaife, M. & Rogers, Y. (1999). Kids as informants: Telling us what we didn't know or confirming what we knew already. In: DRUIN, A. (ed.) *The design of children's technology*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- Sim, G., Macfarlane, S. & Horton, M. (2005). Evaluating usability, fun and learning in educational software for children. *World Conference on Educational Multimedia, Hypermedia and Telecommunications*. Montreal: AACE.
- Zaman, B. & Abeele, V. V. (2010). Laddering with Young Children in User Experience Evaluations: Theoretical Groundings and a Practical Case. IDC, 2010 Barcelona. ACM, 156-165.