Measuring Fun and Enjoyment of Children in a Museum: Evaluating the Smileyometer

F. van der Sluis, E.M.A.G. van Dijk, L.M. Perloy

Human Media Interaction, University of Twente, Enschede, The Netherlands. f.vandersluis@utwente.nl

Introduction

Measuring fun and enjoyment with children is not trivial. Subjective measures are known to suffer from an experimenter effect and often lack detail in their answering. The experimenter effect refers to answering in accordance with the expectations of the experimenter rather than reflecting the opinion of the subject. The lack of detail in answering is due to the oft-found finding (which we will illustrate in this paper as well) that children tend to select the extremes of a scale: either very high or very low, leaving out more fine-grained possibilities. Markopoulos et al. (2008) [1] treat related issues 'satisficing' and 'suggestibility' as important issues to deal with when using survey methods with children. Suggestibility refers to the influence of the experimenter that is generally expected to be more important for children than for adults because of age differences. A respondent that is satisficing gives more or less a superficial answer that appears reasonable without carefully considering the question and the answer options. With children satisficing can easily occur when they find the questions difficult to understand.

With increasing age, children develop increasing skills and ability which explain their different answering tendencies as compared to adults. Piaget noted in his pioneering work the development of thinking skills around the age of 11, where children learned to use abstract concepts and reflect upon them. This includes self-regulation, a skill consisting of self-motivation, attention control, self-monitoring, and self-evaluation [2]. For a subjective measure of fun and enjoyment these skills are pivotal: the ability to monitor oneself and identify these abstract concepts is requisite.

Several attempts have been made to counter the described effects. For example, objective measurements allow (indirect) indicators of fun such as total time spent, total interactions performed, or video recordings. Regarding fun and enjoyment, subjective measures are common and easily applied. Hanna et al. [3] compared the usefulness of several rating methods. Special subjective instruments have been developed for children, such as the Smileyometer or the Visual Analogue Scale (VAS). The Smileyometer (See **Error! Reference source not found.**) uses images of smileys to make the items on the scale more recognizable for the participant. It was designed with the help of children, leading for it to have a non-neutral face at the centre of the scale [4]. In this way, the scale was expected to be more recognisable to children, leading to more fine-grained answers; i.e., less extreme answers. A VAS uses a question format with pictorial representations that children use to identify their feelings or opinions. According to Read et al. [5], VAS scales can be very useful for younger children but when used to elicit opinions about software or hardware products younger children almost always indicated the highest score on the scale. Hence, in this paper we will focus on a popular subjective method for measuring fun with children, the Smileyometer.

The Smileyometer will be reviewed based on two studies we performed in Museon, a science museum in The Hague, The Netherlands. The museum is an educational museum for a broad audience and the main themes are the human and his role in the society, nature, culture and science and technique. Its goals are education, study

Figure 1. The Smileyometer.

Proceedings of Measuring Behavior 2012 (Utrecht, The Netherlands, August 28-31, 2012)
Eds. A.J. Spink, F. Grieco, O.E. Krips, L.W.S. Loijens, L.P.J.J. Noldus, and P.H. Zimmerman

and enjoyment. With over 150.000 visitors each year, of which the majority are children, it is a popular and modern location for fun and learning. As a test location, the museum assures a natural environment, intensifying the need for measurement devices which are reliable and valid in less constraint, non-laboratory settings as well.

Two studies will be reported, both evaluating the enjoyment of children during a quest through the museum. This quest started at a multi-touch table with the selection of topics of interest from the permanent exhibition of Museon.

This selection was used to generate the electronic quest that children answered in the exhibition area. Finally, the quest ended at the multi-touch table with an end game. For both studies an informed consent was given by the parents. The difference between the studies lies in the interfaces used at the multi-touch table, being a direct change of the beginning and the end of the quest and an indirect change on the middle part: the quest changes according to the interactions of the first part. Study One used a prototype whereas Study Two used an evolved version of the system. The focus here, however, will not be on the interfaces but on the Smileyometer, in particular for its experimenter effect and coarse answering.

Study One: Prototype

We used the Smileyometer at the end of the quest through the museum to measure reported enjoyment of the three main parts of the quest: the beginning (initial selection of topics), middle (quest through the permanent exhibition), and end (final game at the multi-touch table). **Error! Reference source not found.** shows the Smileyometer. For each of the three parts of the quest the children used the Smileyometer to answer the question "How much fun was it to do that part?"

Table 1 shows the means and standard deviations (SD) of the first study in which 36 children took part. The answers on the Smileyometer are re-coded to 1 (for awful) until 5 (for brilliant). The results indicate the Smileyometer does not alleviate the mentioned problems: the answers were still almost always very high. Partly this can be explained by the experience actually being fun. Looking at the difference between age groups gives further insights in the causes. Table 1 also shows the descriptive statistics for the younger children (7 to 9) and the older children (10 to 15) separately. Two-tailed t-tests confirmed that for the begin and the middle part, the scores measured with the Smileyometer were significantly higher for children between 7 and 9 compared to children between 10 and 15 (begin part: $t(34)=2.345$, $p= 0.025$, middle part: $t(34)=.0367$, $p=0.004$). However, the final part showed no significant difference. We found significant correlations between the scores on the Smileyometer and age groups for the begin part (Spearman's rho = 0.371, $p=0.026$) and for the middle part (Spearman's rho = 0.425, $p=0.010$). Younger children scored higher. The scores of the final part were not correlated with age. In the begin and middle part the highest score was selected more often by younger children than older children (begin part 58.3% vs. 16.7%, middle part 83.3% vs. 45.8%), no difference was found for the final part (both groups 41.7%). The data indicate younger children have a higher tendency to select the highest ratings, but do select none-extreme scores as well.

Table 1. Enjoyment in a museum quest as measured by the Smileyometer - Descriptive statistics of study 1.

	All			Age 7-9			Age 10-15		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Begin	36	4.11	0.75	12	4.50	0.67	24	3.92	0.72
Middle	36	4.42	0.77	12	4.83	0.39	24	4.21	0.83
End	36	4.11	0.98	12	3.92	1.24	24	4.21	0.83

Table 2. Enjoyment in a museum quest as measured by the Smileyometer - Descriptive statistics of study 2.

	All			Grade 7			Grade 8		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Begin	26	3.77	0.77	16	4.00	0.82	10	3.40	0.52
Middle	48	3.27	1.13	27	3.78	1.05	21	2.62	0.87
End	48	3.73	0.98	27	3.74	0.90	21	3.71	1.10

Study Two: Matured Version

A study similar to Study One was performed at the museum, using a later version of the multi-touch table interfaces at the beginning and at the end of the quest, including improvements based on the evaluation of Study One. This study employed a between-subjects design comparing a personalized quest created by the improved interface versus a normal, fixed, quest without any personalization nor multi-touch interaction at the beginning.

Table 2 shows the descriptive statistics for the Smileyometer of the second study with 48 children. Notice that the number of children that played the initial selection game is roughly half of the total number of participants because of the experimental setup of this study. These results are not extremely high any more. This can partly be explained by the ages of the children. All children were 10-12 years old and were part of two classes (grade 7 and grade 8 in the Dutch school system) of a primary school in The Hague.

We studied age differences again by comparing the results between the class from grade 7 (10-11 years old) and the class from grade 8 (11-12 years old). Table 2 also shows the descriptive statistics per grade level. Even though the difference in age is only one year, again for the begin and the middle part, the scores measured with the Smileyometer were significantly higher for children from grade 7 than for children from grade 8. (results two-tailed t-tests begin part: $t(24)=2.071$, $p= 0.049$; middle part: $t(46)=4.090$, $p<0.001$). Again, no significant difference was found for the final part. We found a strong correlation between the score on the Smileyometer and grade level for the middle part (Spearman's rho = -0.512, $p<0.0005$) and a marginally significant correlation for the begin part (Spearman's rho = -0.375, $p=0.059$). The scores of the final part were not correlated with grade level. Selection of the highest score in the begin and middle part was also very different between children from grade 7 and 8. For the begin part the highest score was chosen by 5 children, all from grade 7 (31.2% vs.0%) In the middle part 8 children, all from grade 7, chose the highest score (29.6% vs. 0%). Again no difference between grade 7 and 8 was found for the final part (22.2% vs. 28.6%). Remarkable here is that the majority of the children who chose the highest score (9 out of 12) were from the group who did not play the initial game and hence played with the multi-touch table for the first time during the end game. The data indicate that even this age difference of one year shows differences in the tendency to select the highest ratings, but at the age of 10-11 years many children do select none-extreme scores as well. There even seems to be a turning point at the age of 11-12 years (just before Dutch children go to secondary school). These children were much more critical in their reported enjoyment.

Conclusion

We reviewed the Smileyometer as used in two field studies, allowing us to compare two different age groups twice: in Study One ages 7-9 against 10-15 and in Study Two ages 10-11 versus 11-12. The results indicate that the Smileyometer, as customized subjective measurement, did not solve the mentioned problems: the reported fun remained high. Both studies indicated age was a factor in the reported fun, such that younger children give higher ratings. The reasons for this effect cannot be proven from the studies, but do point to the problems mentioned in the Introduction: a tendency towards extreme answers and an experimenter effect.

An alternative explanation for the relation between age and reported fun is that younger children simply had more fun. There is no evidence from observation and discussions between the experimenters and the participants to further corroborate this theory, making this an unlikely yet not impossible counter-explanation.

Although the Smileyometer was not able to solve the problems inherent to subjective questionnaires for children, it did partly alleviate the problems. A possible method to further alleviate the problems is by extending the measurement scale from five to, for example, seven items as is often used with semantic differentials. At least for adults, subjective scales tend to become more precise when consisting of more items. For example, this holds even for up to ten items [6].

Summarizing, using pictures was shown to give a reliable scale, and in particular for older children, give fine-grained results. For younger children, however, some problems still remained. Consequently, when using the Smileyometer or similar subjective scales, age should be included in the analyses as a covariant.

Acknowledgements

This work was part of the PuppyIR project, which is supported by a grant of the 7th Framework ICT Programme (FP7-ICT-2007-3) of the European Union.

References

1. Markopoulos, P., Read, J.C., MacFarlane, S., Höysniemi, J. (2008). *Evaluating children's interactive products.* Elsevier Inc., Amsterdam.

2. Ormrod, J.E. (2004). Human Learning. Pearson Education Inc., Englewood Cliffs, NJ, USA.

3. Hanna, L., Neapolitan, D., Risden, K. (2004). Evaluating computer game concepts with children. *Proceedings of Conference on Interaction Design and Children,* 49 - 56.

4. Read, J.C., MacFarlane, S., Casey, C. (2002). Endurability, engagement and expectations: Measuring children's fun. *Proceedings of Conference on Interaction Design and Children*, 189-198.

5. Read, J.C., Gregory, P., MacFarlane, S., McManus, B., Gray, P., Patel, R. (2002). An investigation of participatory design with children - informant, balanced and facilitated design. *Proceedings of Conference on Interaction Design and Children,* 53-64.

6. Van den Broek, E. L., Van der Sluis, F., and Dijkstra, T. (2012). Modeling an acoustic profile for ubiquitous stress assessment. Personal and Ubiquitous Computing **15**(6) (in press).