

Introducing a Pairwise Comparison Scale for UX Evaluations with Preschoolers

Bieke Zaman

CUO/IBBT, KULeuven, Parkstraat 45 bus 3605, 3000 Leuven, Belgium
{Bieke.Zaman}@soc.kuleuven.be

Abstract. This paper describes the development and validation of a pairwise comparison scale for user experience (UX) evaluations with preschoolers. More particularly, the dimensionality, reliability and validity of the scale are discussed. The results of three experiments among almost 170 preschoolers show that user experience cannot be measured quantitatively as a multi-dimensional construct. In contrast, preschoolers' UX should be measured directly as a one-dimensional higher order construct. This one-dimensional scale encompassing five general items proved to be internally consistent and valid providing evidence of a solid theory-based instrument to measure UX with preschoolers.

Keywords: UX evaluation, pairwise comparison scale, preschoolers

1 Introduction

Making decisions on the development and launch of new technologies has become very difficult. Nowadays, complex product characteristics matter as unique selling points to distinguish from competitors. Companies have to come up with innovative ideas that generate a noticeable user experience that totally fits today's users' unfulfilled, and often unspecified dreams and wishes. These user experiences are often hard to reveal, and even harder to measure. Nonetheless, the launch of a new or improved digital product totally depends on whether it leads to a better user experience than the previous versions or competitor's products. The importance of good benchmarking measurement tools for decision makers can thus no longer be neglected.

Standardized questioning based on *closed-ended scales* is necessary in order to benchmark and compare user experiences across products. For each product or user tested, the conditions –items and underlying scale– should be similar. This way, closed-ended scales allow measuring quantitative differences in product preferences based on user experiences.

Because of the ungraspable or intangible character of user experiences, we cannot analyze user experience as an independent measure. Instead, comparable or relative measures are needed. *Relative scaling* reveals a motivated domination of some objects over others with respect to a common attribute [1]. In our case, one technology might be preferred over another with respect to the user experience it brings to the user. Each relative UX score depends on the score on the same attribute

UX of another object. This contrasts physical scaling which scores concrete attributes such as length, temperature or weight independently [1]. The strength of asking respondents to make relative comparisons and judgments, rather than absolute judgments has first been theorized by Thurstone in 1927 [2].

In 2004, Hanna et al. [3] suggested the potential of *pairwise comparisons* for the evaluation of digital technologies with children. As far as we know, no research project ever took that challenge. In general, quantitative methods that rely on relative and closed-ended scaling for UX evaluations with young children rarely exist. Most methods for preschoolers are qualitative, and the existing quantitative methods are often not appropriate to work with this age group who cannot yet read or write. In this paper, however, we will suggest a pairwise comparison scale for UX evaluations that is adapted to work with preschoolers.

2 The UX Scale for Preschoolers

In the next paragraphs, we will report on the results of three experiments in which a UX scale for preschoolers was developed and tested. The UX scale is part of the *This-or-That method*, a mixed-method approach to benchmark children's user experiences with technologies [4]. Chronologically, the This-or-That method equates a within-subject experiment consisting of four phases in which children are invited individually to judge preferences on user experiences of two technologies. During the first phase, children can explore both technologies. Secondly, a quantitative survey questionnaire (based on the UX scale) is verbally administered to let children judge preferences between both conditions. The answers are then validated qualitatively through a short probing interview. Finally, their preferences are also validated through a behavioural choice: at the end of the experiment, we allowed children to play one of the two conditions again as 'a reward for good participation' ('free play option').

2.1 The Dimensionality of the UX Scale for Preschoolers

In [5], we described how literature review on the one hand, and empirical qualitative data on the other hand revealed a preliminary five component classification of the UX construct for preschoolers: 1) challenge & control, 2) fantasy, 3) creative and constructive expressions, 4) social experiences, 5) body and senses. These categories have been repeatedly reported as explanatory factors for what children positively experience and thus like in technology. We then hypothesized that these categories are correlated and together measure the underlying construct of user experience. This classification should make from UX a construct that can be measured multi-dimensionally.

Three out of five components were measured by five specific questions each and tested in an experiment with 36 preschoolers ($N_{girls}:17$, $N_{boys}:19$ - $Age:65$ months, $SD_{age}:8$ months). In the experiment, we added five more general questions to control whether UX can also directly be measured on a one-dimensional level. Because of the low *internal consistency* of the three specific subscales related to the three specific UX components (Chronbach's alphas with values lower than .6), we decided to

further test the one-dimensional UX (sub)scale with preschoolers. The Chronbach's alpha for the one-dimensional UX subscale indicated a good reliability: .882 ($M=7.91$, $SD=2.050$). The high internal consistency of this last overall UX scale was confirmed in a second experiment among 113 preschoolers ($N_{boys}:56$, $N_{girls}:57$, $M_{age}:58,39$ months or 5 years, $SD_{age}:14$ months) resulting in a Chronbach's alpha of .797 ($M=6,147$, $SD=1,569$).

Results of *principal component analysis* confirm that UX cannot be measured multi-dimensionally with preschoolers. We had to refute our hypothesis that for each specific component of our UX construct the corresponding variables would correlate.

We can conclude that although the multi-dimensionality of UX can be measured in a qualitative way [7], its multi-dimensionality cannot be measured quantitatively with such a young age group. UX should instead be measured directly as a one-dimensional higher order construct. This one-dimensional scale measures UX via the following five general items: 1) "Show me which product was most fun?" 2) "Show me which product would you like to receive as a gift?" 3) "Show me which product would you like to take home?" 4) "Show me which product would you like to play again?" 5) "Show me which product you found a little bit stupid?"

2.2 Convergent Validity, Criterion Validity and Scale Validity

The *convergent validity* of the general UX scale was assessed by comparing its scores to a free play option. In both experiments, we found a significant correlation between the first free play moment and the scale ($r=.570$, $N=33$ for the first experiment, and $r=.581$, $N=111$ for the second experiment, both Kendall's tau at the $p<.01$ level).

As for the scale's *criterion validity*, many research papers do report on the relationship between usability and hedonic qualities [e.g. 6,7]. Our results are in line with previous studies. We indeed found a significant correlation between the results on our UX scale and the usability of the product. More particularly, the results of our first experiment showed significant correlations between the UX scale and both the completion time with both interfaces as well as the successful completion of the game. More details on these correlations are discussed in [8]. In sum, the correlations between the UX scale and usability in our first experiment suggest a good criterion validity of our scale.

A third experiment was set up with 18 children ($M_{age}: 4.3$ years, $N_{boys}:9$, $N_{girls}:9$) to test the accuracy of preschoolers' responses (*scale validity*). In this third experiment, children were asked to judge two identical games, so that there was no difference between both conditions, except from the actual UX. The researcher did not mention anything about the fact that they had to play and judge one game twice.

The results of the third experiment show that children only had difficulties or doubts to state preferences as an answer to the first question; after that first question they could answer more easily. To our surprise, the overall product preference expressed via the UX scale corresponded well with the free play option (Kendall's tau $r=.659$, $N=13$, $p<.01$). The significant correlation between children's answers and their behavioural preference refutes the idea that children were inventing answers to please the adult. In contrast, the results assume that children discerned real differences

in user experiences between the alternatives and answered consistently. The qualitative interview conducted after the questionnaire, confirmed that children were judging the experiences they had with the game rather than the game characteristics in isolation. Most of their answers reflected a more positive UX towards the second condition than the first, which was most likely due to the fact that they were more experienced in the second condition and therefore could reach higher goals as well.

3 Conclusion

In this paper, we aimed at filling the gap on quantitative methods that allow for product benchmarking and UX evaluations with preschoolers. More particularly, we found that pairwise comparison techniques lead to reliable answers from these young children who cannot yet read or write. The paper introduced a valid five-item pairwise comparison scale that allows for relative measurement of UX and comparisons across two products in terms of preferences. The scale is perfectly suitable for prototyping to decide on which alternative designers should continue. In such a design context, the quantitative phase should be complemented with qualitative research to get insight into the reasons for UX preference in order to recommend design improvements. In follow-up papers, we will document more on this mixed-method This-or-That approach -including the UX scale and qualitative interview techniques- and how/why it is exactly so appropriate to use with preschoolers. We will also go more deeply into the possibilities that our quantitative and pairwise comparison scale of UX provide.

References

1. Saaty, T L.: Relative Measurement and its Generalization in Decision Making: Why Pairwise Comparisons are Central in Mathematics for the Measurement of Intangible Factors-The Analytic Hierarchy/Network Process. RACSAM, 102 (2), 251-318 (2008)
2. Thurstone, L.L.: A law of comparative judgement. Psychological Review. 101(2), 266-27 (1994)
3. Hanna, L., Neapolitan, D., Ridsen, K.: Evaluating computer game concepts with children. In: IDC2004, ACM Press, Maryland (2004)
4. Zaman, B., Abeele, V.: How to Measure the Likeability of Tangible Interaction with Preschoolers. In Proc. CHI Nederland, pp. 57-59. Infotec Nederland BV Woerden (2007)
5. Zaman, B., Abeele, V.: Towards a Likeability Framework that meets Child-Computer Interaction & Communication Sciences. In Proc. of IDC, pp. 1-8 (2007).
6. Hassenzahl, M., Tractinsky, N.: User experience – a research agenda. B&IT, 25(2), 91-97. UK: Taylor & Francis (2006)
7. Tractinsky, N.: Aesthetics and apparent usability: Empirically assessing cultural and methodological issues. In Proc. of CHI, pp. 115-122 (1997)
8. Abeele, V., Zaman, B., Vanden Abeele, M. (2008, October). The Unlikeability of a Cuddle Toy Interface: An experimental study of preschoolers' likeability and usability of a 3D game played with a cuddle toy versus a keyboard. Book Chapter in Lecture Notes in Computer Science Volume 5294/2008, 118-131 Berlin: Springer: (2008)