
Paper Ladder: A Rating Scale to Collect Children's Opinion in User Studies

Cristina Maria Sylla

University of Minho
Braga, Portugal
sylla@engagelab.org

Eva Irene Brooks

Aalborg University
Aalborg, Denmark
eb@learning.aau.dk

Ahmed Sabbir Arif

University of California, Merced
Merced, California, USA
asarif@ucmerced.edu

Elena Márquez Segura

University of California, Santa Cruz
Santa Cruz, California, USA
elena.marquez@ucsc.edu

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

MobileHCI '17, September 4–7, 2017, Vienna, Austria
© 2017 Copyright is held by the owner/author(s).
ACM ISBN 978-1-4503-5075-4/17/09.
<https://doi.org/10.1145/3098279.3122139>

Abstract

We present Paper Ladder, a low-cost, light-weight, and paper-based variant of the Sticky Ladder rating scale to collect children's product preferences in user studies. This paper version revamps the original, effective but arguably impractical evaluation method by making it more accessible to researchers. In this paper, we show its potential with preliminary results from a pilot study, where 45 grade-2 children used a Paper Ladder to rate their preference for different types of correctness feedback in a math app. Results suggest that Paper Ladder could be an effective and valuable instrument for conducting user studies with young children.

Author Keywords

Children; survey methods; evaluation; user studies.

ACM Classification Keywords

H.5.2. User Interfaces: Evaluation/methodology.

Introduction

Survey methods are an important evaluation instrument to collect opinions, thoughts, and perspectives of technology usage and experience from a target group, independently of its demographic characteristics [3]. Researchers in the field of Child-Computer Interaction (CCI) have emphasized the importance of asking children directly about their preference when designing



Figure 1. A completed Sticky Ladder, from Airey et al. [1].

and evaluating interactive products for them since children perceive the world differently from adults and have different needs and preferences [2,4,7,10,16].

However, collecting young children’s opinion in user studies is challenging. Depending on their age and cognitive development, children may have difficulties in understanding the survey questions and to clearly communicate their thoughts verbally. Further, “*children usually have little control over the opportunities that may arise for them to express their views to adults*” [10]. In addition, children may aspire to please adults when expressing their views, influencing their responses [1,4,8,9]. Thus, special care must be taken when selecting a survey method for a user study involving children.

Airey et al. [1] attempted to address these issues with the Sticky Ladder rating scale. With this scale, children express their preferences for different products by physically sticking tangible Velcro items representing the products onto a small fabric ladder (Figure 1). This method was designed based on the idea that children are better at expressing their preferences by physically manipulating tangible objects instead of articulating thoughts through words. A user study with ten 4-6 years old children suggested that young children can understand and use the Sticky Ladder rating scale without any major difficulties. Nevertheless, this rating scale is rarely applied (we found only one instance of use [19]), possibly because of the time, effort, and cost of preparing the logistics of the method. To address this, we propose a lightweight and more accessible paper-version of the scale, named Paper Ladder (Figure 2), which we made publicly available for download at <http://www.asarif.com/resources/PaperLadder>.

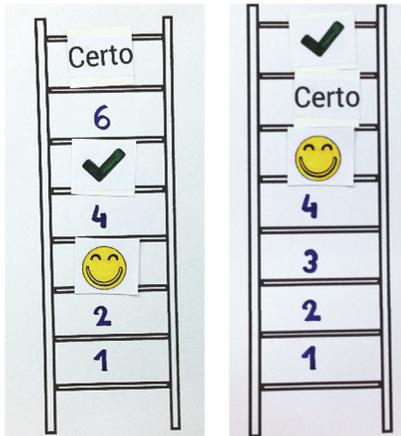


Figure 2. Two completed Paper Ladders.

Related Work

To enable both children and adults to participate in the design process of novel technologies for children, Druin [6] developed the cooperative inquiry approach. It was applied in a lab setting with design teams that included adults and 7-11-year-old children to gather field data, initiate ideas and tests, and develop new prototypes. She concluded that the cooperative inquiry approach “*can lead to exciting results in the development of new technologies and design-centered learning*”. However, this approach requires both adults and children to observe, take notes, and interact with child users, which may be too demanding for young children. Hunleth [11] used a similar approach in an ethnography study, where children between the ages of 8 and 12 were asked to draw, tape-record, and participate in focus group discussions to express their thoughts. She cautioned about the risks of using child-oriented approaches without critical reflection, which include a limited access, awareness, and appreciation of the children’s experiences and knowledge, and a reinforced pre-conceived image of the child.

Survey methods have been used with children to collect information on user interface design, particularly to evaluate children’s preferences [1,20], explore the “appeal” and “fun” of a product [16], and to understand children’s mental models and processes [4,17]. A widely-used approach for adapting a reliable rating scale (particularly the Likert scale) for children involves using text-based categorical response options. For example, the Visual Analogue Scale (VAS) is a graphical rating scale displayed as a continuum line between two extremes and the Numeric VAS uses numerical categorical responses [14]. Several variations of these scales attempt to reduce the number of response

choices or craft the wording to match the vocabulary and cognitive skills of children [15].

Some also use smileys as response choices. For instance, Smileyometer is a smiley version of a five-point Likert scale that measures different dimensions of fun with 5-10 years old children [16,20]. These types of methods replace complex language with concepts that are familiar to children, such as emoticons, and could be as effective as the Likert scale. However, evidence suggests that young children often have difficulties with differentiating between the different smiley faces expressing certain degrees of preference, such as "especially happy" and "slightly happy" [1,16]. In addition, young children do not usually use ratings in the middle of a scale [5]. Multiple investigations suggested that young children tend to bias their answers toward the extreme positive side [1,4,5,16,18]. Further, 6-10 years old children have a tendency to score most products as "brilliant" [16,20], presumably because they usually engage in dichotomous thinking and focus on the extremes of a scale. They also find it difficult to make fine-grained discriminations of emotions [5]. These issues could affect both the classical Likert scale as well as its expressive smiley variants.

Relevantly, in a comprehensive study involving over 300 children aged 9-11 years, Hall et al. [8] found out that children usually use two points of the smiley scale: "positive" and "very positive". Hence, to encourage children to make full use of the scale, they developed the Five Degrees of Happiness scale that only presents positive emotions. A user study revealed that this method helped 9-11 years old children to use the full scale to communicate their judgements [8].

Other adaptations of survey methods attempt to bypass children's difficulty to verbally express ideas by asking them to "act" their preferences instead. Examples of such methods are the This-or-That [20] and the Sticky Ladder [1]. The This-or-That method carries out pairwise comparisons of products by asking children to show their preferences by pointing to a preferred product among two. The Sticky Ladder allows the comparison between multiple products simultaneously and rate them according to various factors, such as "enjoyment" and "liking" [1]. Both methods avoid difficult instructions and allow children to express their views without having to verbalize them. However, we find the Sticky Ladder more useful because: first, it could be used to compare multiple products at the same time. Then, it enables the use of different evaluation criteria. Last, the physical properties of the scale are valuable from a situated cognition perspective since it allows children to think through their actions, such as when (re)considering their preferences as they (re)organize the products. This physical rearrangement could as well be used to prompt children to think aloud and discuss their actions and thinking.

Paper Ladder

The Paper Ladder deploys the Sticky Ladder rating scale [1] and its items (i.e., the products under evaluation) on printed cutouts instead of a stuffed ladder and items that can be attached with Velcro.

The Paper Ladder used in our evaluation was created by drawing a ladder on an A4 sheet of paper representing a seven-point Likert scale. The product pieces were created by cutting small pieces of paper (50×50 mm) with the pictures of the three different

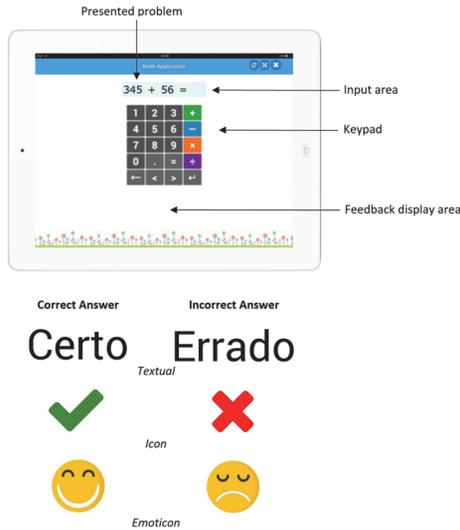


Figure 3. The device and the application used in the study. Below, the different types of visual feedback provided on a correct or an incorrect answer.



Figure 4. A child participating in the user study.

types of correctness feedback provided by a math app for children (Figure 2 and Figure 5).

This paper-based variant of the Sticky Ladder is low-cost and more easily deployable than its original counterpart. Another advantage of this variant is that it is easier to create multiple representations of the same product. This can allow researchers to emphasize different products' features, and potentially evoke and evaluate different aspects of the product. Finally, from a situated cognition perspective, the non-adhesive nature of the pieces representing different products and/or features is advantageous since it could enable children to perceive their choices as more provisional, offering them the possibility to (re)consider their choices as they move/slide the items around on the ladder. These advantages could potentially encourage researchers to resort to this method for gathering children's opinions in user studies.

User Study

In this section, we report preliminary findings from a pilot study that explored the following aspects.

- Ease of use of the Paper Ladder from the perspective of the children and the researcher.
- Data variance: whether and how children's responses varied over the whole scale.
- Advantages of using manipulation of the cutout items to evaluate the products instead of verbalizing preferences.
- Insights into children's thinking process while they are comparing and manipulating items.

We used a Paper Ladder to assess children's preference of different types of correctness feedback used in a

custom math app (Figure 3). The study was carried with three grade-2 classes from a local school, each consisting of 15 children (in a total of 45). Their average age was 7.06 years (SD = 0.25). The first class had 8 females and 7 males, the second had 6 females and 9 males, and the third class had 8 females and 7 males. All children received consents from their parents to participate in the study. We used three Apple iPad 3 Wi-Fi tablets for the study. The tablets were placed on three tables distant from each other to ensure that children worked on their own.

Group Class	Session1 Day 1	Session2 Day 2	Session3 Day 3
1	Emoticon	Textual	Icon
2	Icon	Emoticon	Textual
3	Textual	Icon	Emoticon

Difficulty Levels Easy ■ Moderate ■ Hard ■

Table 1. Design of the user study.

The custom app presented children with drill questions for addition and subtraction. The drill questions were categorized into easy, moderate, and hard difficulty levels. It displayed one problem at a time, and prompted children to enter the answer using a keypad. Upon each entry, it displayed either textual, icon, or emoticon correctness feedback.

We used a mixed design for the study. Each class used the three feedback types with alternating difficulty levels in three sessions that expanded over three days to reduce any potential effects of skill transference. Hence, the independent variables or conditions were

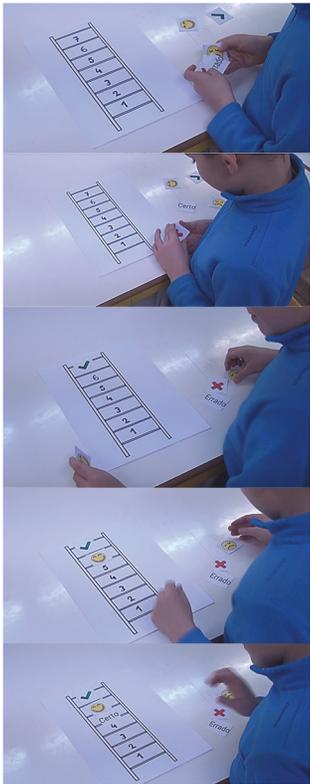


Figure 5. Sequence showing a child placing tangible cards on a Paper Ladder.

feedback type \times difficulty level (Table 1). We used this design to explore if different difficulty levels affected children's preference of the correctness feedback.

The study was conducted during regular class-hours. We started by demonstrating the app and the feedback types, and then asked children to try their best to solve the problem (Figure 4). However, we assured them that it was alright to make mistakes as we were investigating their preference for the correctness feedback types not how well they performed solving the drill questions.

Using the Paper Ladder

Following the completion of the study, we asked children to rate the feedback types using a Paper Ladder. The evaluation was carried at school with one child at a time to avoid any potential bias due to mutual influence. We started by showing the ladder and the ladder pieces to the child, explaining how each piece represented a feedback type in the app. We then demonstrated how positioning the pieces on different steps of the ladder represented preference, such as upper steps for the preferred choice and bottom steps for the less preferred choices. Once the child understood the functionality of the scale, (s)he was asked to rate the examined feedback types.

We avoided guiding children with questions. However, the researcher formulated the following prompts to help two children (out of 45) that needed assistance: "Which method did you like the most?", "Which method did you like the second?", and "Which method did you like the least?". These questions were alternated for the two children.

Observations and Discussion

In this section, we present our observations and discuss the results of the study.

Ease of Use

Almost all children (43 out of 45) understood the Paper Ladder without any major difficulties and seemed to be comfortable using it. They were all able to map the pieces to the feedback types effortlessly. Further, judging from their engagement, they seemed to enjoy using it.

Manipulation of Items

The pieces representing the three feedback types were easy for the children to handle. Most of them started by placing their most preferred feedback type on the top of the ladder, and then used the bottom steps for the least preferred ones. Some children also started with their least preferred feedback types. Several children changed their ratings by moving and rearranging the pieces on the ladder. The fact that the pieces were not adhesive may have enabled children to perceive their choices as more provisional. This would explain their behavior of pondering over (presumably reconsidering) their choices as they moved/slid the items around on the Paper Ladder.

Access to Children's Thinking Process

Children were silent and focused while using the Paper Ladder. However, we think that the way they looked at, held, and manipulated the product cards could potentially provide insights regarding their thought processes. For example, children frequently reordered the pieces on the ladder, which could indicate that they were re-evaluating their preferences.

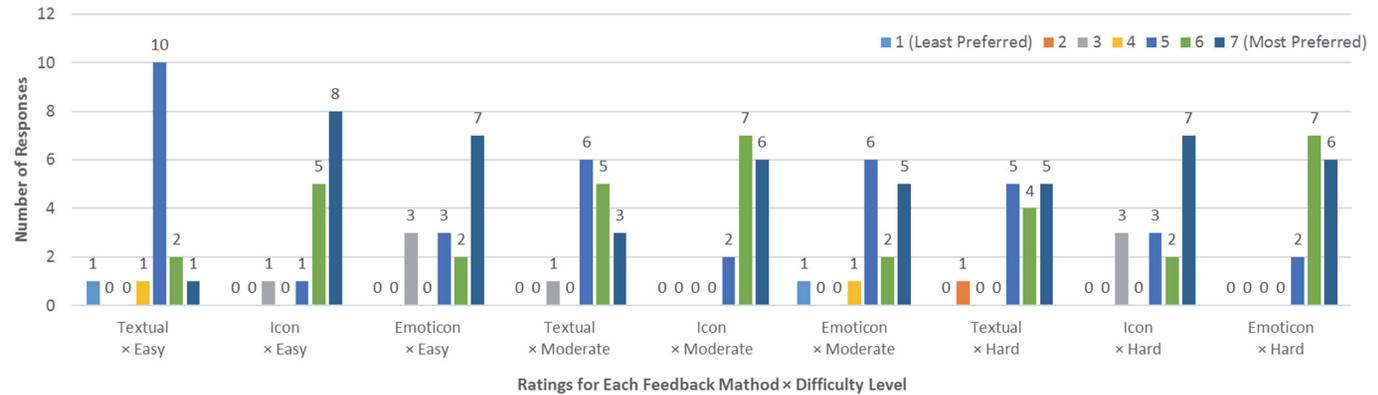


Figure 6. Children’s rating of all examined correctness feedback.

Data Variance

As mentioned earlier, the three main shortcomings of the existing scales are that young children often find it difficult to differentiate between the ratings expressing the different levels of similar responses [1], they do not use ratings in the middle of the scale [8], and they bias their responses toward the extreme positive side [1,4,5]. Figure 6 displays the variation of children’s ratings across the scale, roughly 1.5%, 1%, 6%, 1.5%, 28%, 26.5%, 35.5% ratings of 1, 2, 3, 4, 5, 6, 7 on a seven-point scale, respectively. This indicates that some children were able to differentiate between the different levels and to express different degrees of preference. Besides, about 36% of all ratings were in the middle scale (i.e., from 3 to 5), and only about one third of all responses (35.5%) were at the extreme positive side. This suggests that the Paper Ladder could be an effective method for collecting children’s opinion in user studies. However, further investigation is necessary to fully explore this possibility.

The study showed that the Paper Ladder was easy to use for the researchers since the customization of the scale only took some minutes. In addition, it was also easy to use for the children. As anticipated, the children mapped the paper pieces to the feedback types easily. Moreover, the physicality of the tool made it easier for them to visualize their ratings and rearrange them by sliding the cards on the ladder before finalizing their choices. This procedure gave us an insight into children’s cognitive process while comparing the items. Children seemed to feel comfortable and enjoy using the scale throughout the study. One possible reason for this could be the manipulation of tangible objects, which facilitates cognitive processes [13] and reduces cognitive load [12].

Regarding the quantitative results, the method was successful in collecting data with a relatively high degree of variability (Figure 6). This suggests that the Paper Ladder encourages the use of points at the

middle of the scale, providing an effective method for collecting children's opinion in empirical studies.

Conclusion

We presented the Paper Ladder, an easy-to-craft and customize low-cost variant of the Sticky Ladder rating scale that could be used to collect children's product preferences in user studies. We explored its effectiveness in a pilot study, where children seemed to feel comfortable and enjoy using the scale. Results suggest that the tool can encourage children to use all scale points, thus providing a promising method for collecting young children's opinion.

Limitations and Future Work

Results of the pilot suggest that the Paper Ladder can be a useful tool in collecting young children's opinion in user studies. It, however, does not tell us whether it performs better compared to Sticky Ladder and other popular rating scales for children. In the future, we will conduct a comparative study to fully explore this.

One limitation of the pilot study is that it used an unorthodox study design where the conditions were not properly balanced. The three user groups did not experience the same difficulty levels with all feedback types. This may have influenced their ratings of the feedback types. We will address this in a future study by properly (counter)balancing the conditions. Further, the evaluated feedback types were a within-subjects factors in this pilot study. To investigate if the scale works well with other study designs, we will use a between-subjects design in the future. We will also use a more structured method for dividing children into different groups based on their verbal and cognitive skills. Additionally, will measure the understandability,

usability, enjoyability, and likability of the Paper Ladder rating scale, and the benefits of the non-adhesive item cards from a situated cognition perspective.

Finally, the pilot study only evaluated visual items (i.e., different types of visual feedback). In the future, we will investigate if the scale is also useful when children are asked to rate preferences for non-visual items or activities.

Acknowledgements

We would like to thank the students and staff of the school EB1 Gualtar-Braga in Portugal for participating in the study. Cristina Sylla is financed by the Portuguese Foundation for Science and Technology (FCT) and co-financed by the European Regional Development Fund (ERDF) through the Operational Programme for Competitiveness and Internationalization—COMPETE 2020, within the Research Centre on Child Studies of the University of Minho (CIEC) POCI-01-0145 FEDER-007562 with the Postdoctoral Grant: SFRH/BPD/111891/2015.

References

1. Sharon Airey, Lydia Plowman, Daniel Connolly, and Rosemary Luckin. 2002. Rating children's enjoyment of toys, games and media. In *Proceeding of the 3rd World Congress of the International Toy Research Association on Toys, Games and Media*, 1–7.
2. Ahmed Sabbir Arif and Cristina Sylla. 2013. A comparative evaluation of touch and pen gestures for adult and child users. In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13*, 392–395. <https://doi.org/10.1145/2485760.2485804>
3. Ida Bark, Asbjørn Følstad, and Jan Gulliksen. 2006. Use and usefulness of HCI methods: results from an exploratory study among nordic HCI practitioners. In *People and Computers XIX — The Bigger Picture*.

- Springer London, London, 201–217.
https://doi.org/10.1007/1-84628-249-7_13
4. Alice Bell. 2007. Designing and testing questionnaires for children. *Journal of Research in Nursing* 12, 5: 461–469. <https://doi.org/10.1177/17449871079616>
 5. Christine T. Chambers. 2002. Developmental differences in children's use of rating scales. *Journal of Pediatric Psychology* 27, 1: 27–36. <https://doi.org/10.1093/jpepsy/27.1.27>
 6. Allison Druin. 1999. Cooperative inquiry: developing new technologies for children with children. In *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit - CHI '99*, 592–599. <https://doi.org/10.1145/302979.303166>
 7. Allison Druin. 2002. The role of children in the design of new technology. *Behaviour & Information Technology* 21, 1: 1–25. <https://doi.org/10.1080/01449290110108659>
 8. Lynne Hall, Colette Hume, and Sarah Tazzyman. 2016. Five degrees of happiness: effective smiley face Likert Scales for evaluating with children. In *Proceedings of the The 15th International Conference on Interaction Design and Children - IDC '16*, 311–321. <https://doi.org/10.1145/2930674.2930719>
 9. Libby Hanna, Kirsten Ridsen, and Kirsten Alexander. 1997. Guidelines for usability testing with children. *interactions* 4, 5: 9–14. <https://doi.org/10.1145/264044.264045>
 10. Malcolm Hill. 2006. Children's voices on ways of having a voice: children's and young people's perspectives on methods used in research and consultation. *Childhood* 13, 1: 69–89. <https://doi.org/10.1177/0907568206059972>
 11. Jean Hunleth. 2011. Beyond on or with: questioning power dynamics and knowledge production in "child-oriented" research methodology. *Childhood* 18, 1: 81–93. <https://doi.org/10.1177/0907568210371234>
 12. David Kirsh. 1995. Complementary strategies: why we use our hands when we think. In *Proceedings of the 17th Annual Conference of the Cognitive Science Society - CogSci '95*, 212–217.
 13. Scott R. Klemmer, Björn Hartmann, and Leila Takayama. 2006. How bodies matter: five themes for interaction design. In *Proceedings of the 6th ACM conference on Designing Interactive systems - DIS '06*, 140. <https://doi.org/10.1145/1142405.1142429>
 14. H van Laerhoven, H J van der Zaag-Loonen, and B H F Derkx. 2004. A comparison of Likert scale and visual analogue scales as response options in children's questionnaires. *Acta paediatrica* 93, 6: 830–835. <https://doi.org/10.1111/j.1651-2227.2004.tb03026.x>
 15. David Mellor and Kathleen A. Moore. 2014. The use of Likert scales with children. *Journal of Pediatric Psychology* 39, 3: 369–379. <https://doi.org/10.1093/jpepsy/jst079>
 16. Janet C. Read and Stuart MacFarlane. 2002. Endurability, engagement and expectations: measuring children's fun. *Interaction Design and Children* 2: 1–23. <https://doi.org/10.1.1.100.9319>
 17. Janet C Read, Stuart MacFarlane, and Chris Casey. 2003. What's going on?: discovering what children understand about handwriting recognition interfaces. In *Proceeding of the 2003 conference on Interaction design and children - IDC '03*, 135. <https://doi.org/10.1145/953536.953555>
 18. Christine M Rubie-Davies and John A C Hattie. 2012. The dangers of extreme positive responses in Likert scales administered to young children. *The International Journal of Educational and Psychological Assessment* 11: 75–89.
 19. Annika Wolff, Paul Mulholland, Zdenek Zdrahal, and Richard Joiner. 2007. Re-using digital narrative content in interactive games. *International Journal of Human-Computer Studies* 65, 3: 244–272. <https://doi.org/10.1016/j.ijhcs.2006.10.003>
 20. Bieke Zaman, Vero Vanden Abeele, and Dirk De Grooff. 2013. Measuring product liking in preschool children: an evaluation of the Smileyometer and This or That methods. *International Journal of Child-Computer Interaction* 1, 2: 61–70. <https://doi.org/10.1016/j.ijcci.2012.12.001>