Evaluation Methods for Children's Technology

Yaël M. Richter-Symanek
Universität Bremen
Bremen, Germany
yrichter@uni-bremen.de

1. ABSTRACT

As child-computer interaction has become a field of its own detached from the general field of human-computer interaction with more attention to the fact how different the requirements for children's technology are, it has become equally important to evaluate those technologies sensitive to the differences in approach and specific challenges when working with children.

This paper discusses several common evaluation methods as well as evaluation frameworks aimed at categorising and giving an overview to researchers. Each method is presented with its advantages and disadvantages regarding a conduction with children as participants and supplemented by notable research on the topic. Discussed evaluation frameworks include those by Markopoulos and Bekker and the PLU-E framework by McKnight and Read. As an introduction to the field, the paper gives a brief summary of the specific challenges studies using children as participants face.

Author Keywords

Evaluation methods; evaluation frameworks; usability; children; human-computer interaction; child-computer interaction.

2. INTRODUCTION

In the field of Human-Computer Interaction *Usability* is a common term used to describe not only the *efficiency* (i.e. How much time does it take for a user to complete a given task using the technology?), the *learnability* (i.e. How long does it take a user to learn how to use the technology?), the *memorability* (i.e. How well can a user remember how to use the technology?) and the *safety* (i.e. How well is a user protected from making mistakes in the use of the technology and how extensive and useful is the technology's offered assistance?) but also the *level of satisfaction* users experience while using it [14]. These factors can often determine whether a technology will be used or rejected in consideration to each factor's importance given the specific situation and application environment. As a result, reflecting on a technology's *Usability* has become crucial for companies as well as research groups working on all kinds of technology. The variation in focus in regard to context and product requires an equally wide range of methods to choose from to evaluate. Because evaluating is what one has to do to know for sure whether set up goals are met.

For instance, an acceptance test, where human performance is measured quantitatively to see if it falls within an acceptable criteria (for example, a certain time limit to complete a given task, a targeted error rate or a level of relative satisfaction), can be handy − especially for technology designed for learning purposes − to prove that the examined technology outperforms existing technologies [6]. When designing a vocabulary trainer one could try to show that students studying for a test with the help of this particular software perform better than their classmates who studied using a classical approach or another technology. Not only could such an evaluation validate and verify that the designers indeed constructed the right software in the right way, one could also make use of the results later on to convince students, parents or teachers that they should use the product. This may apply not only to companies trying to sell a product, but also to government institutions revising school curricula or researchers proposing new educational approaches and is not limited to normal day-to-day technology used by adults, as this example shows. In fact, with the growing market of computerized toys it has become more important than ever to make those meaningful, enjoyable and useful for their young users.

However, for children's technology the circumstances and target audience usually require a technique different from the ones commonly used for evaluating adult's technology and experts on those are not always familiar with the specific challenges connected to evaluating children's technology the same way many university lecturers would struggle teaching kindergarten kids. These are the reasons why the field of child-computer interaction has established itself as a whole new area of research in the last few years.

Upon understanding that evaluating technology designed for children comes with a whole new set of challenges whose solving one has to become familiar with before attempting a successful evaluation, having a look at the specific problems prior to discussing possible solutions seems a sensible idea.

Specific Challenges in Evaluating Children's Technology

For a successful evaluation, children have to fit in a role in other contexts usually filled by adult test participants where they are observed while using a technology, so the

Yaël M. Richter-Symanek, part of the seminar 03-ME-801.99a on Child Computer Interaction (WiSe 2017/2018), Universität Bremen.

researchers conducting the evaluation get an insight on occurring issues and can gather results to fix possible bugs. As there is only so much an attentive observer can learn from an outside perspective, evaluation frequently includes some kind of interview or questionnaire. Some more structured than others depending on the product and the background and nature of the participants at hand. Consequently, to fill this role kids will have to understand given questions, recall relevant information for answering these − for example, about the technology and if they liked or disliked something about it − and form an appropriate response that not only includes their opinion, wishes and needs but at the same time is understandable for the adults conducting the evaluation.

In effect to kids taking on roles initially designed for adult participants, kids will struggle to perform the given task as their brains work differently and hence the problems that can occur differ greatly from those of adults in evaluation, although, some turn out to be surprisingly alike (we will come to this when discussing specific methods). Common problems broken down to basic principles of the human mind include − but are not limited to − issues of articulation, attention span (it is already quite hard to keep an adult focused for a longer period of time and everyone who has worked with children knows how easily distracted they are by more interesting topics or just basic needs) and memory [3] [12].

Of course, these problems of the human nature aside, it has also proven to be a challenge in itself to get access to enough children of a certain age and target group and that for a sufficient period of time, which sometimes turns out to be the first and greatest challenge for this kind of project. Due to busy school schedules and the fact that adult researchers usually will have better access to adult participants than to younger children, studies on evaluating with children tend to be rather short, which further promotes the difficulty of inexperienced researchers together with the already mentioned limited resources resulting from the restricted time children can concentrate and are available [13].

Finally and above all, children will need some kind of motivation or reward for participating. The best way to motivate them is, certainly, to make participating fun. Yet this is often more easily said than done and requires similar sensitivity of adult researchers as the choice of method and measured factors in general. Still, for choosing an appropriate method one can nowadays draw from an ever getting wider amount of research and experience or completely rely on evaluation frameworks constructed by experts that will give advice on which method suits a particular technology and group of subjects best, which promise to make the process of conducting an evaluation easier and more feasible.

This paper will give an overview on both in its following sections − some common evaluation methods and their pros and cons as well as proposed frameworks for their categorisation as a system to choose a method. As a means of example, it will refer to the tiptoi books by Ravensburger at various locations to show how a method may be applied.

Tiptoi books are books that can be navigated with the help of an electronic pen that recognises points on the pages and plays audio accordingly − for example, conversations, music or typical sounds related to the displayed scene or object depending on the chosen settings. Furthermore, users can play searching or guessing games using book and pen. The publishing house offers books on different topics such as countries, maths, the alphabet, types of animals or historical periods [19].

3. METHODS

Naturally, researchers ask themselves which methods they can use to evaluate. Sometimes the decision process is complicated by the fact that different methods overlap, are partly similar or have fuzzy dividing lines because they are usually used combined. Having a look at different types and categories can help to distinguish which one of those to choose for further examinations of its requirements and advantages and finding the one best suited for the technology in question.

Field Studies

To begin with, it is important to differentiate between evaluation methods that take place in a controlled lab environment and those conducted in the field meaning in an environment natural to the users and comparable to the one the technology will be used in later.

Field studies vary a lot in terms of interaction between researcher and user. Some classical methods are purely observational while others require researchers to become part of the group they design for. Typical methods include interviews (especially dynamic ones without predefined questions) and ethnographic research [4]. They have become more popular in recent years because of the specific challenges mobile devices bring that cannot be replicated in the lab as you can take a mobile device anywhere, hence, the technology has to be flexible and work the same way in many different environments and situations which makes it often necessary to take these factors into account when testing for bugs concerning hardware or usability. Also, they may be a good second step to traditional usability testing in lab conditions after some bugs have been smoothed out [8], although they may be equally helpful in the early stages of designing to get first ideas on how the technology would be used, in what context and what it needs to be able to do [4].

When it comes to children, this can mean visiting a school or kindergarten and observing the kids while they play to get to know the context or − when some developmental stage of prototype or the technology itself already exists −

how they approach and handle the technology when it is given to them to get an idea of how intuitive it is or if it can keep children interested for a longer period of time. For children, participating in a field study can be more comfortable than doing so in a lab environment − especially for shyer kids − as the environment is more natural to them and will reassure them in feeling safe and talking more in contrast to visiting a new environment in a lab study [12].

Nevertheless, the lack of a controlled environment can make evaluation through field studies harder due to changes in variables and missing reproducibility. Moreover, the amount of data - for example, collected by mobile tracking devices - may be overwhelming and misdirect and confuse in regard to what is of importance and what is not.

While a familiar environment can provide children with a feeling of safety when needed, it might also further increase concentration problems due to a less structured setting and possible distractions, for example, by peers, more interesting objects or the events around them [12].

For the tiptoi, one could imagine conducting a field study in a kindergarten or school. For example, for the tiptoi "Deutsch: 1. Klasse" book it may be helpful to observe what and how children learn at that age in a real German class, what the material is and what interests the children the most. However, challenges can occur regarding permissions from teachers and parents (they will naturally ask what their kids will gain from the study, how they will profit) and not all kids might be interested in the topic. In this context it could be easier to have a look at existing school books for the first grade instead or to consult teachers who have got experience teaching children that age as the context of use is not that important for this specific technology and scenario.

Generally, field studies are favourable when the context of use is especially relevant for the product. In nearly all other cases lab studies should be preferred due to superior control and cheaper conduction. Still, exceptions may prove the rule when there are important reasons to do so.

Informal evaluations

A method that directly involves possible users (unlike those that do not which this paper will discuss later in this section) is getting informal feedback on an idea or early stage of a product. These informal evaluations are often conducted using a low-fidelity prototype.

Using low-fidelity prototypes to get feedback from children is particularly helpful for certain kinds of technology, as Rick et al. discuss in connection with tabletop interfaces as children already have experience with physical components, do not have to use their less developed finer motor skills in most cases and can concentrate on the task at hand rather than on possible problems concerning the prototype's interface [16].

These early feedbacks are cheap to conduct and require little planning due to their informal nature [8]. Furthermore, the use of low-fidelity prototypes can provide a flexible solution where changes can easily be made and children find themselves in a playful environment that comes naturally to them. Especially, when combined with a collaborative approach where they can work together on tasks and communicate freely while the researcher merely observes and guides when necessary [16]. However, low-fidelity prototypes sometimes come with affordances that are not part of the technology itself. This can lead to children using the prototype differently from how it was intended. For example, moving or lifting an item that represents an object on a display [8].

The tiptoi books are by now a technology well-established in Ravensburger's publishing concept, however, a few years back at the beginning of that journey and testing the general idea of a book that is expanded by audio content through the touch of a pen is something where the conduction of an informal evaluation using a low-fidelity prototype could prove to be helpful. Such a prototype might consist of a normal picture book or drawings and a stick-like object, while audio feedback on each touch is provided through recordings played by a supervisor hidden from the children. The advantage here would certainly be that it could be conducted cheaply, without lengthy preparations and could possibly provide insightful feedback on the general concept of tiptoi books. In this, informal evaluations seem preferable to field studies.

In short, informal evaluations give a low-cost opportunity for input from real users − in our case children − and new ideas when used with an open mind, willingness to listen as well as flexibility in regard to the outcome and possible changes.

Usability testing

Where informal evaluations usually give no instructions to participants on how to give feedback and work more as a method of observation, classical usability testing provides formal instructions while users are asked to complete certain tasks. Normally conducted in a lab environment, variables, conditions as well as the collected data are easily controlled and standardised. Common ways to collect feedback are think-aloud protocols in combination with audio or video recording for subsequent analyses. This makes for a good method to measure performance in terms of accuracy and efficiency and furthermore, often allows insight into the user's thought processes [8].

However, with formality some issues arise and are to be expected when dealing with children. One crucial example are the problems kids have with the think-aloud method as it requires not only a certain level of articulation but also a general motivation to verbalise thoughts. Some children find themselves inhibited − especially in front of adult researchers [12] − and often need to be encouraged to talk about their impressions and opinions.

A lot of different approaches have been attempted to overcome this problem. Some include pairing children up for co-discovery sessions (where they work in pairs to complete a task) [12] or peer tutoring (where one child who has already completed the task teaches another how to use the technology) [9], others making use of some more uncommon tools such as a social-robot instead of an adult supervisor to free the young participants of their inhibition when thinking-aloud [5]. Some studies tried encouraging feedback through think-aloud retrospection where children would watch recordings of themselves working on a task [17].

In the course of this, different usability studies came to different conclusions suggesting that the success of each specific method depends highly on the product, the tasks, the participants used and their age and, after all, on the particular way it is conducted. Nevertheless, some tendencies seem to exists, making it likely that formal usability testing is better used with older children and teenagers due to their superior articulation and concentration together with their better developed and trained ability to follow instructions. Another tendency supports the assumption that children who work in pairs provide better feedback if they like and know each other [1] [7].

The difficulties concerning usability testing and younger children suggest to consider carefully whether to use it on a type of product such as the tiptoi books as those are clearly targeted at children of kindergarten and primary school age. However, usability testing could still prove to be useful when it comes to the technology surrounding the tiptoi where, as an example, audio material first has to be downloaded by parents or teachers to be accessed from the tiptoi pen. This hardly falls into the area of evaluations with children and therefore underlies the principles of adult usability testing which are not part of this paper. Still, especially for more technical problems controlled and reproducible results are needed and formal methods such as usability testing appropriate.

Questionnaires

As a method of collecting self-reported data, questionnaires are often used in combination with another evaluation method and in most cases after participants worked with a technology to gather their opinion and preferences. However, questionnaires can be used on their own and without prior examinations of a technology [8] and provide a useful procedure to collect data on general preferences or a certain context of use that is cheaper than many other evaluation methods and organises data in a neat way which makes analysing it easier and faster. This is especially true when using standardised questionnaires.

Using questionnaires with children normally makes it necessary for them to be able to read and write to understand what is asked of them and fill in given questions, but researchers can bypass this by assisting children, reading questions to them or using questionnaires that can be understood through symbols (this unfortunately voids the advantage that all participants are asked the same questions in the same manner they have over, for example, interviews). One example for the latter type of using symbols is the *Smileyometer*, which asks children to rate a technology with the help of smilies. However, different studies concluded that children tended to rate all the given technologies overly positive. Better results were achieved using methods that ask children to rank the technologies (*Fun Sorter*) or decide which one of two they would like to use again (*This or That* method) [11] [18]. Still, the findings suggest that questionnaires need to be designed carefully for the targeted age group and have to make sure children's understanding of the questions does not interfere with the results.

However, one could imagine using online questionnaires with older kids who are already actively using technology in their day-to-day life as studies have shown that using paper-based questionnaires compared to online-based ones does not have an influence on the results and children's understanding [10]. Especially in likewise cases, questionnaires appear to be superior to interviews because no attending staff is needed to conduct questionnaires children can answer on their own.

Interviews

As a technique of gaining information, interviews vary in regard to formalisation and are adaptable to the desired approach as well as the type of data that one wants to collect, but are usually used in combination with another method for a wider picture of the application context or user-acceptance [8]. Generally, interviews fall into three categories − namely structured, semi-structured and dynamic or rather informal ones. For structured or semi-structured approaches standardised frameworks can be found where the interview follows predefined questions and/or a system is given to analyse quantitative data outcomes. Dynamic interviews, on the other hand, offer a more flexible method of getting feedback compared to classical questionnaires.

When it comes to children, interviews may face similar challenges to those of informal evaluations, where children need to be encouraged to verbalise their thoughts in an appropriate and coherent way − especially younger kids.

Studies found that probing specific events with the help of video recall (showing recordings of themselves using the technology to them) to assist children in remembering relevant information or specific information was successful as well as giving children the opportunity to demonstrate their thoughts during an interview by providing them with the technology in question. Price and Jewitt (2013) came to the conclusion that both approaches gave different types of results which suggests that choosing carefully between the different methods is crucial to obtaining specific information [15] [8].

Expert reviews

In contrast to methods involving users in a direct way, expert reviews offer one that is usually cheaper and quicker to conduct without the need to find children of a targeted age and application context. Expert reviews are performed by experts on the field of usability for children, often with the help of certain principal heuristics or checklists to identify possible bugs and problems [8].

The perks of consulting experts are certainly that feedback comes quickly and in a form easily understandable to product designers and researchers. Often even accompanied by advice on how to fix identified bugs, expert reviews make for a good first step to eliminate major bugs prior to further evaluation by the means of other methods.

However, as expert reviews do not require involving children in the process, they deal with a set of problems different from the ones described in this paper's introduction. The main focus here is on how thoroughly experts can find bugs and if the predicted problems turn out to be problems when children use the technology.

Baauw et al. (2006) discuss this for a method of structured expert evaluation called SEEM and came to the conclusion that expert reviews were able to identify most problems but that certain ones indeed turned out not to be of importance when it came to children's usage of the technology [8] [2].

Regarding the tiptoi technology, expert reviews may be especially helpful when it comes to the general usage of the pen asking questions such as "Is the log-in procedure for the pen at the beginning of each book understandable?", "How clearly is the navigation on each page structured?", "Does the user know his or her options at every moment?" or "Does the user always know the status of the pen and which mode is active?" and finding possible usability bugs.

4. EVALUATION FRAMEWORKS

Motivated by the lack of a general systematic comparison of different methods several researchers have constructed frameworks to fill this gap and help with the choosing of method and participants for evaluations with children in the past few years [12]. This paper will discuss two different frameworks in the following section − the one proposed by Markopoulos and Bekker and the PLU-E framework McKnight and Read created in 2011.

Markopoulos and Bekker (2003)

Focusing on comparing usability testing methods for children the framework establishes three dimensions for organising studies and their results.

The first dimension formulates criteria to generally describe a usability testing method. One being *robustness* that describes how adaptable a method is to different contexts of use (for example, field versus lab environments). *Reliability* assesses the usual definition of this term as whether the method achieves the same results when performed for the same product in the same manner repeatedly. Whether found problems are actual usability problems is described by the factor of *validity* supplemented by *thoroughness and problem counting* to assess the actual number of problems uncovered. The last criteria of *efficiency* measures the amount of resources used in relation to the outcome.

The second dimension aims at assessing general key elements of the studies to provide researchers with a basic knowledge of how to apply the described method. It focuses on parameters such as the number of participants, the context in which it was conducted, the tasks given to the participants, the technique used (for example, interview, questionnaires, video recordings) as well as the procedure in general.

The third and last dimension focuses on the children who participate as test subjects and how their individual characteristics and abilities influence the test's outcome. Markopoulos and Bekker list many different criteria here ranging from general points, such as gender, to cognitive and motor skills to factors closer related to an individual's personality.

They point out that many previous comparisons of usability testing methods already use parts of the first two dimensions described by them but that research should focus more on the third dimension as the characteristics of the participants used in relation to the target group a product aims at and the conduction of usability testing in general seem to be of great importance [12]. Therefore, using the framework proposed by them appears to be a useful method to bring a new viewpoint to choosing usability testing methods for research and could be a valuable addition before deciding on a group of participants.

McKnight and Read (2011) - PLU-E Framework

Like Markopoulos and Bekker, the creators of the PLU-E (*Playing Learning and Using for Evaluation*) framework motivate their efforts with the need for a clear approach to evaluation and the need to prevent weak evaluations before they are conducted. The model was based on the PLU model first proposed by Read in 2005 that describes how children interact with technology categorised as three types of interaction where children take the role of either players, learners or users. In regard to which role children fill or rather what the technology aims at (for example, being a functional product for attending to a task or being an educational tool to teach a child certain parts of a material) different aspects should be primarily evaluated. On this foundation, McKnight and Read construct their PLU-E framework to guide the decision process for an evaluation method and formulate step-by-step instructions.

These steps consists of choosing a main purpose for the technology with regard to especially important components, choosing (a) target group(s), agreeing on a PLU weighting that represents the product best (to what extend is the

product aimed at playing, learning and using in percent points), deciding on when in the process evaluations are needed as well as planning the evaluations based on the first steps and general constraints such as time and resources.

Using the PLU outline it is possible to categorise evaluation methods into the different categories and determine which aspects of those they test best. However, not all methods will be easily categorised as for testing one aspect only, but a general tendency can be determined and individual decisions be made.

McKnight and Read already propose a sorting of several common evaluation methods according to the PLU model in their paper, including the previously mentioned *This or That* and *SEEM* method [13]. This might serve as a foundation for further categorisations or choosing a method with the help of the PLU-E framework. Still, designers and researchers should be aware that the PLU-E framework is merely aimed at being an assisting tool in the process of choosing an approach to evaluation and is not intended to serve without the addition of careful weighting and considerations [13]. Moreover, categorising into only three general terms does not leave much room for grey areas and products out of the norm, so any variations here will have to be handled individually and with regard to their details. Nonetheless, McKnight and Read propose a useful categorisation and provide further material for all design processes to which their method of classification can easily be applied.

All in all, frameworks can provide a useful systematic approach to choosing a evaluation method, when one understands the way they categorise and what aspects are the most important ones for the technology at hand. In general, every guideline and rule should only be followed when understanding the reasoning behind its construction.

5. DISCUSSION

After examining different methods of evaluation as well as on proposed systems for their categorisation, several important conclusions stand out. Firstly, one can see how diverse evaluation methods truly are and that each one has its own advantages and disadvantages. That said, which one to choose can be a difficult decision that is not easily made and can be facilitated only to a certain extend by considering what other studies used in the past as every product and pool of participants require an individual approach with regard to their very own characteristics and targeted context of use.

However, to some extent a system and general heuristics have to be established as the field of child-computer interaction gets more important in the context of usability and a growing industry of technology for children. McKnight and Read as well as Markopoulos and Bekker propose two different frameworks focusing on different aspects, providing different information from different viewpoints on the matter. Which one of those to choose from − or if to use the help of frameworks at all − remains a decision as individual to product and researchers as the final one concerning the evaluation method. Nevertheless, evaluation frameworks give helpful advice to evaluators who have little or no experience when it comes to conducting evaluations with and for children.

6. CONCLUSION

This paper discussed different approaches to evaluation for children's technology and aimed at giving a structured overview on a variation of some more common methods and classification systems to provide researchers with a sound foundation for conducting their own evaluations and understanding what is discussed in academic work on this topic. Still, the real effectiveness of all the mentioned methods and frameworks in different contexts needs to be further examined by more studies as the field of child-computer interaction grows in a society that is subject to a far-reaching technological change. We do not know for sure how much contact children will have with technology in ten or twenty years but with a view to current developments it seems more important than ever to make the technology children use in their day to day life − be it for fun or educational purposes − meaningful and usable in a way that will leave them with curiosity and self-confidence in their role as users. Evaluating carefully while being attentive to different product, target and age groups is certainly a step in the right direction.

7. REFERENCES

1. Als, Benedikte S., Janne J. Jensen, and Mikael B. Skov. "Comparison of think-aloud and constructive interaction in usability testing with children." *Proceedings of the 2005 conference on Interaction design and children*. ACM, 2005.

2. Baauw, Ester, Mathilde M. Bekker, and Panos Markopoulos. "Assessing the applicability of the structured expert evaluation method (SEEM) for a wider age group." *Proceedings of the 2006 conference on Interaction design and children*. ACM, 2006.

3. Egloff, Tammie Hutto. "Edutainment: a case study of interactive cd-rom playsets." *Computers in Entertainment (CIE)* 2.1 (2004): 13-13.

4. Farrell, Susan. “Field Studies.” *Nielsen Norman Group*, Nielsen Norman Group, 23 Oct. 2016, www.nngroup.com/articles/field-studies/.

5. Fransen, Sjef, and Panos Markopoulos. "Let robots do the talking." *International Journal of Arts and Technology* 5.2-4 (2012): 293-318.

6. Greenberg, Saul, and Bill Buxton. "Usability evaluation considered harmful (some of the time)." *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2008.

7. Hanna, Libby, Denise Neapolitan, and Kirsten Risden. "Evaluating computer game concepts with children."

Proceedings of the 2004 conference on Interaction design and children: building a community. ACM, 2004.

8. Hourcade, Juan Pablo. "Child-computer interaction." *Self* (2015).
9. Höysniemi, Johanna, Perttu Hämäläinen, and Laura Turkki. "Using peer tutoring in evaluating the usability of a physically interactive computer game with children." *Interacting with computers* 15.2 (2003): 203-225.
10. Kano, Akiyo, and Janet Read. "Interchangeability of computer and paper based questionnaires in gathering computer experience data from young children." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2012.
11. MacFarlane, Stuart, Gavin Sim, and Matthew Horton. "Assessing usability and fun in educational software." *Proceedings of the 2005 conference on Interaction design and children*. ACM, 2005.
12. Markopoulos, Panos, and Mathilde Bekker. "On the assessment of usability testing methods for children." *Interacting with computers* 15.2 (2003): 227-243.
13. McKnight, Lorna, and Janet C. Read. "Plu-e: a proposed framework for planning and conducting evaluation studies with children." *Proceedings of the 25th BCS Conference on Human-Computer Interaction*. British Computer Society, 2011.
14. Nielsen, Jakob. *Usability engineering*. Elsevier, 1994.
15. Price, Sara, and Carey Jewitt. "Interview approaches to researching embodiment." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2013.
16. Rick, Jochen, et al. "Lo-fi prototyping to design interactive-tabletop applications for children." *Proceedings of the 9th International Conference on Interaction Design and Children*. ACM, 2010.
17. Van Kesteren, Ilse EH, et al. "Assessing usability evaluation methods on their effectiveness to elicit verbal comments from children subjects." *Proceedings of the 2003 conference on Interaction design and children*. ACM, 2003.
18. Zaman, Bieke, Vero Vanden Abeele, and Dirk De Grooff. "Measuring product liking in preschool children: An evaluation of the Smileyometer and This or That methods." *International Journal of Child-Computer Interaction* 1.2 (2013): 61-70.
19. *tiptoi® Konzept*, Ravensburger, www.tiptoi.com/de/tiptoi-konzept/.