Jun 25th, 9:00 AM

Usability testing with children: History of best practices, comparison of methods & gaps in literature

Andria Banker
University of Minnesota-Twin Cities, United States of America

Carlye Lauff
University of Minnesota-Twin Cities, United States of America

Citation

Banker, A., and Lauff, C. (2022) Usability testing with children: History of best practices, comparison of methods & gaps in literature, in Lockton, D., Lenzi, S., Hekkert, P., Oak, A., Sádaba, J., Lloyd, P. (eds.), *DRS2022: Bilbao*, 25 June - 3 July, Bilbao, Spain. https://doi.org/10.21606/drs.2022.646

DRS 2022
Bilbao 25th June — 3rd July

Design Research Society

Usability testing with children: History of best practices, comparison of methods & gaps in literature

Andria Banker, Carlye Lauff*

Product Design Program, University of Minnesota – Twin Cities, USA

*corresponding e-mail: carlye@umn.edu

doi.org/10.21606/drs.2022.646

Abstract: This paper is a systematized literature review centered on the best practices of usability testing with children, when they are the target end user in product development. The paper begins with a brief history of usability testing with children during the prototyping stage of product development. Following, the methodology guiding this literature review is described. Then, guidelines for usability testing with children provided by past research are outlined in chronological order, documented to show the evolution of changes or improvements in the practices over time. Additionally, most of the approaches to usability (as part of evaluative research) have been identified and compared between one another with a variety of factors. In conclusion, directions for further research are suggested based on current unanswered questions in the field of prototype usability testing with children, such as considerations for longitudinal vs. cross-sectional testing, physical vs. digital product testing, and age range of children.

Keywords: children; prototyping; usability testing; evaluative research

1. Introduction and background

1.1 User-centered design process

User-centered design focuses on creating ideal products and experiences for the anticipated user. In this approach, oftentimes end users participate in aspects of the design process, where they can express their opinions on the design (Sanders, 2002). There are multiple design process models; one model is the UK Design Council double diamond design process, where there are four basic phases: discover, define, develop, and deliver (Ball, 2019). This paper focuses on the develop and deliver, when prototypes are being built and tested (Gustafsson, 2019). Prototypes are “physical or digital embodiments of critical elements of the intended design” (Lauff et al., 2018). It is important to test prototypes with the intended user to make revisions in design and/or to solve usability problems.

1.2 The role of children in the design process

Allison Druin (2002) developed the categorization of roles children can hold in the design process: user, tester, informant, and design partner. The design partner spends the most

time in product development, whereas the user spends the least time. Children as users is the first role that they had in design. Providing feedback on products as a user helps guide future products, but not the current product in development. The tester role is to evaluate prototypes to make further iterations of prototypes until the design goals are met. Children as informants help provide information and feedback during various stages, including ideation. Design partners are treated as equals to designers and are available throughout the entire design process (Druin, 2002). This paper focuses on the role of children as the tester. This decision was made based on practical concerns of industry: time allotted for research and the necessity to get feedback from users before the product is manufactured.

1.3 Evaluative research and usability testing

Evaluative research is a type of human research that uses methods and techniques to measure and evaluate a product, service, offering, or intervention. Typically, this research is done after prototypes are developed. Usability testing is a subset of evaluative research, in that usability testing refers to evaluating a product or service by testing it with representative users. In the context of this work, those representative users are children. The intention of usability testing is to uncover both positive and negative aspects of the usability interactions with the intended product (Dumas & Fox, 2007). Markopoulos et al. (2008) states that usability testing focuses on how the design of a product will support users in performing their tasks, with prototypes focusing on interaction and the functionality of the product. For example, researchers developing ScreenPlay, an interactive toy for waiting spaces in hospitals, tested the product with 11 children in order to verify the product reached the goal of creating a positive experience for children without engaging in contact with the product's surface. Using the usability testing methods of questionnaires, focus groups, and observations, the researchers were able to identify aspects of the toy that needed changed for it to be more enjoyable for the children (Biddiss et al., 2013). It is the goal of this literature review to first document the best practices for usability testing with children and then compare usability testing methods across several factors, including the age range of children, location for the testing, and need for moderator.

1.4 Importance of testing with children

For products designed for children, it is paramount that the usability testing occurs with children, since they are the intended users. In the United States, by law a child is defined as anyone under the age 19 (Dependents 2 | Internal Revenue Service, 2021). In psychology, the definition of children is further broken down into the cognitive-developmental stages determined by Jean Piaget (Markopoulos, Read, et al., 2008). These stages are the sensorimotor stage (0 months-2 years), the preoperative stage (2-7 years), the stage of concrete operations (7-11 years), the stage of formal operations (11+ years) (Coelho & Fernandes, 2013). The age group that the product is developed for should be the group that is tested.

While it is essential to involve representative end users in usability testing, this has not been the established practice for products designed for children (Markopoulos et al., 2008).

Historically, the typical evaluation method for children's products was to ask experts on children, experts on usability testing, or to ask the children's parents or teachers their opinions on the product (Markopoulos et al., 2008 & Druin, 2002). All three of these categories of people do not include children, who are the intended user.

Children and adults have significant differences, ranging from physical features, cognitive development, and amount of personal and life experiences. Children also differ in their culture, norms, complexities, likes, dislikes, curiosities, and needs. Designers may assume they know what children like or dislike because of having been a child before or interacting with children. While this may provide some background knowledge, it can create biases and blind spots when designing products. The circumstances and environment surrounding designers' childhoods are very different from present-day children's experiences (Druin, 2002). Snitker (2021) coined the term kids' experience, or "KX", and explains the difference between children and adult researchers as the autonomy of person and the ability to adapt to circumstances. Adults can be their own person and have control over their choices; they are also able to adapt their behavior to different positions, whether it be researcher or respondent. Snitker (2021) explains that when doing design research with children, the researcher must build a bridge between the adult and child's world. Children will provide honest feedback, but it is up to the adult researchers to document and interpret that feedback.

1.5 Guidelines and biases in testing

Guidelines for evaluative research are established to minimize biases and ensure consistency across testing populations. Bias can cause research studies to be prejudiced and, specifically in usability studies, can create an inaccurate representation of future product use. Cognitive bias, the tendency to subconsciously change one's choices based on someone or something else's comments or behavior, decreases the validity of a study's outcome (Natesan et al., 2016). The human brain reacts to stimuli around it very quickly in a series of shortlists, called cognitive biases. These are rules that the brain has set up to be able to create an answer quickly, like a "jerk-reaction" or a reflex (Snitker, 2021). Cognitive biases cannot be eliminated in a study, but they can be minimized. To do this, a researcher must be aware of what biases can occur and ways to correct them (Natesan et al., 2016.). This leads to the use of guidelines, or best practices. Guidelines are a series of procedures that have been demonstrated either by research or experience to show reliable results.

There are more than 180 cognitive biases that have been documented (Manoogian, 2016). Understanding these biases has been helpful in many fields, such as business (Schwenk, 1986), marketing (Kienzler, 2018), and even fire safety (Kinsey et al., 2019). Cognitive biases occur for product testers, regardless of age. The same biases can occur in children as well as adults, but they can also be very specific to children. These biases can occur either between the researcher and the child or between the child and the product. For example, the researcher can bias the child by asking questions in a positive light, such as, "Was the product enjoyable?". An example of an instance where children have a bias toward a product is a

case in which the prototype breaks and the child then holds a negative attitude toward the idea of the product because of the negative experience of it breaking. Cognitive bias directly relating to children during a usability study can be found during the set-up of the study, the introductions given by the moderator, the duration of the test, and after the test is over.

An example of a bias during the set-up is in planning the location. If the location is comfortable to children, such as if the furniture fits the child, then they will be more likely to communicate their thoughts and feelings because they will feel less isolated than in a space specific to adults (Snitker, 2021). Since the location directly affects the choices the child will make, the goal of the researcher is to choose a location that enables feedback from the child while allowing the child to stay focused on the product. Priming is another cognitive bias in which the person experiencing the bias will be expecting what they are primed to expect (Natesan et al., 2016). For example, if a researcher is overly exuberant about the product, the child will already be expecting something "great". While the end goal of a product designer is to have a product the child will enjoy, the goal for usability testing is to find the usability problems. It should be expected that there will be some negative feedback when testing products, and for the product designer to then take that feedback and improve the product. Additionally, children may change their behavior depending on the behavior of the evaluator during testing. If a moderator is taking notes, the child may interpret that action as the child herself being tested, this may, in turn, make the child feel pressure and be more nervous to do or say things (Ellis et al., 2008). A bias after testing that could occur is asking questions that are not clear or concise. For instance, if a researcher asks children a question in which they include a metaphor that the child does not understand, then the children often interpret the meaning of metaphors quite literally and lose understanding of the question altogether, unable to give an answer.

In summary, there are many cognitive biases that can occur during usability testing with children. While it is nearly impossible to remove all biases, there are best practices to consider when preparing, executing, and concluding a usability test.

1.6 Purpose of review

In product design, it is common practice to consider different targeted user populations in your research. More recently, there is a growing field of research studying children as the end user for products, like toys (Markopoulos, 2003). The purpose of this review is to provide an overview of guidelines and best practices for carrying out usability studies with children that are currently found in literature. A chronologically organized assessment of the literature will be presented to display the foundational practices and changes to the guidelines for usability testing with children over time. Then, evaluative research methods that are typically used during usability testing will be compared based on existing comparison studies presented in the literature. This synthesis will assist in enabling practitioners of usability testing, especially those new to the field, to be prepared for children as participants and will demonstrate where there are needs or gaps in the current literature.

2. Literature review research methodology

This research paper is a systematized literature review centered on usability testing guidelines and methods, specifically for products for children. As a systematized review, this research attempt to include elements of systematic review process, with synthesis in a narrative and tabular format to help identify what is known through new lenses, recommendations for practice, and some areas for future research. As this is a small, but growing, field of study, there is limited literature on the topic of usability testing with children. In the timespan of three months, 105 articles of literature were found through the process of searching and screening titles using databases of peer-reviewed sources. The databases used were Google Scholar and the University of Minnesota Library citations database. Keywords used to find literature were 'children,' 'usability testing,' 'evaluative testing,' 'product testing,' 'play-testing,' and 'toy testing.' Further literature was found by reviewing accomplished authors on the topic and identifying their co-authors and collaborators, and then further analyzing the works of those researchers. After screening the abstracts, literature was included if it referred to the process of usability testing with children or guidelines of conducting product research with children. Literature about the general topics of usability testing and of children were included to provide a basic framework leading to the specific topic of evaluative testing with children. The total number of papers included was seventy, provided in the references section. The lead graduate researcher primarily conducted the literature review with support from a faculty researcher.

In analyzing the literature, the identified papers were read, and pertinent information was extracted from each paper. The data gathered was then reorganized and synthesized to illuminate opportunities for exploration. These opportunities led us to comparing guidelines and methods for usability testing for children written by researchers in the field. In the organization of the paper, the general guidelines for using any usability testing method for children is presented along with an overview of why it was necessary to include such guidelines. The holistic guidelines created are a structured comparison of four guidelines that are intended for practitioners of usability testing with children. They are organized by the stage of the study: set-up, introduction, during the testing, and after the testing. The synthesis also includes a review and comparison of the types of usability methods used, and then the age of children involved, moderator involvement during the testing, and location where the testing occurs. Additionally, the research approach to each paper was identified, whether based on personal experience, literature review, research studies, or a combination. Lastly, the types of products were identified: whether digital, analog, or a combination of digital and analog. It is the goal of this paper to condense all this valuable information into one place, both to help practitioners and to help researchers identify the gaps in the field.

3. Guidelines for usability testing with children

3.1 Chosen sources and evolution of guidelines

As discussed in Section 1.5, there are many cognitive biases that can occur during usability testing with children. While it is nearly impossible to remove all biases, there are best practices to consider when preparing, executing, and concluding a usability test. We identified four of the most prominent literature sources for guidelines for usability testing; these sources are shown in Table 1 along with what research was conducted to developed these guidelines (personal experiences, literature review research studies). This classification was determined by the source's self-proclamation or by thoroughly reviewing the literature. Then, in Table 2, these four sources were evaluated and simplified into basic guidelines for working with children during usability testing. The goal of these guidelines is to be a quick reference for practitioners, regardless of the evaluative research method or location, as well as a way to see the evolution of guidelines from 1999 through 2021. The table begins with Guidelines for Usability Testing with Children by Hanna et al. (1999), because these are the guidelines in which many modern sources in the study of usability testing with children reference as the first comprehensive guidelines on the subject (Barendregt & Bekker, 2003). Then, in chronological order, the works of Barendregt and Bekker (2003), Markopoulos et al. (2008), and Snitker (2021) are each documented into a simplified, yet comprehensive set of best practices that have evolved over time. Each source builds on the source before it, showing how through time, more is known in the field and there is a need for additional rules or suggestions to create a better practice of usability testing with children.

Table 2 is divided into four sections of the evaluation process: set-up, introduction, during the testing, and after the testing. In these subsections, the guidelines were evaluated to determine if they would have a direct relation to the child and the biases that could occur with relation to the child during the testing. For example, in Snitker's guidelines (2021), he provided information to guide the stakeholders to have a comprehensive understanding of the project before investing time, money, and resources into the testing. This is an important practice, but it does not directly affect the children involved with the project. Because this paper is focused on working with children, the guidelines were analyzed through that lens.

Table 1. Comparison of basis of guidelines for usability testing with children*

	Personal Experience	Literature Review	Research Studies
Hanna et al. (1999)	✔		
Barendregt & Bekker (2003)			✔
Markopoulos et al. (2008)	✔	✔	✔
Snitker (2021)	✔		✔

* The basis of guidelines is how the authors gathered their information to form their guidelines. Personal experience refers to their own experience of usability testing of products. Literature review refers to documenting what others have confirmed in previous literature on the topic of usability testing with children. Research studies involve experiments or comparison studies the authors have completed.

Table 2. Evolution of Usability Testing Guidelines based on the Phase of Research

Author, Title, Publication Year	Hanna et al. *Guidelines for Usability Testing with Children.* 1999	Barendregt & Bekker *Guidelines for User Testing with Children* 2003	Markopoulos et al. *Evaluating Children's Interactive Products: Principles and Practices for Interaction Designers* 2008	Snitker *User Research with Kids: How to Effectively Conduct Research with Participants Aged 3-16* 2021
Phase of Research				
Set-up and Planning	The setting should be comfortable but not distracting. Secondary technology should be familiar to children. Observation equipment (video cameras or recorders) are unobtrusive. Only schedule the children for an hour each, giving breaks when the children start to run out of energy. Give yourself breaks between children in order to refresh. Switch the order of tasks for different children. Screen the children beforehand to make sure they have enough	Five or more children should be tested. Include children that will experience many problems and be willing to verbalize. Use pictures to explain what information you are seeking from the children. Keep pictures around for children to refer to during the test.	Be aware of the location you will be testing in, visit beforehand. Be prepared for any disruptions or unplanned changes. Have everything you need before testing takes place. Inform adults in the vicinity of the test when there will be children nearby. If the study takes place over several days, be sure to keep the time of the testing constant. Be aware that children will be more tired after lunch or dinner and be more alert in the mornings.	Use skill level to describe the respondents. When developing the test-script, keep the participant at the forefront. Plan the testing for an accurate time when the product would be in actual use if manufactured and in the child's known settings. Make sure children are familiar with their surroundings by inviting them a day earlier to get a tour or going to a location the child already knows.

	understanding to be able to use the product. It is not a good idea to use your own or colleagues' children for testing.		Children have more energy on bad weather days when they are unable to play outside. Be prepared to deal with the findings of the study in the light of this. All children in the study should be given a chance to take part in order for the evaluation to be fair. Plan to have at least two adults present. Have a backup of any technologies or products needed. Make sure that the designed tasks do not demand a higher skill level than the children have. Plan the testing space so the children will be comfortable to sit quietly for an extended period of time. The pilot test should include 2-3 children that are the same age as those being tested. Plan to test the youngest end of the targeted age group. If a number of participants is needed, know that there may be no-shows and design that into the procedure.	
Introductions	Use small talk with children to establish a baseline relationship.	Introductions should be natural and relaxed for the children to feel relaxed.	Give instructions as statements rather than questions. Speak to each child individually.	Do not put emphasis on any aspect of the child, it is important that the child follows their natural interests.

	In order to explain confidentiality agreements in terms the children will understand, tell the children the designs are "top secret." Parents must sign the agreement. Have a script prepared. Emphasize that the children are not being tested, the product is being tested. For motivation, emphasize the importance of the role. Introduce prototype so that children know what to expect. Explain the setup in detail. Explain the roles of moderators, one-way mirrors, any devices the children can see.		Instructions should be clear and unambiguous and consistent for all children.	Instructions need to provide a rich understanding but not overwhelm.
During Testing	Give children a simple task or toy not related to the product in order for them to relax and get into the testing headspace. Make sure children understand the tasks, restate if necessary. Redirect children's questions with questions of your own. To direct toward a new task, do not ask if the child wants to, but instead, say "Now I need you to…" or "Let's do this…"	Respond to children's remarks in a concise manner. Have a list of appropriate ways to respond to children at hand during testing. Be willing to improvise when working with children. Before helping the children, encourage them to try a bit longer. Free play should be allowed before task play.	Never blame the children if something goes wrong. Give parents a questionnaire on their observations if they are also in the observation room. Keep a timer handy in order to stay on schedule. Children should not be expected to stay on one task longer than 20 minutes. If there is a waiting period before a child is tested, the waiting room should be	If there are multiple products, give the child enough time with each in order for them to be able to recall the product later if interviewing. Break the session into 15-minute intervals. The moderator should be neutral in dress tone, and body language. The moderator should match the child's energy levels.

	If children become distracted from the product, gently remind them to pay attention. If a child is uninterested in the product, draw them in by pretending you need help doing it. Make sure to provide breaks when necessary. If reading words or numbers is necessary for the product's use and the child is struggling with that part, you can read for them. Be sure not to provide answers for them. Keep motivation going by offering generic positive feedback. An adult should be in the room with the child. It is better to have a moderator, that is trained not to interfere, than a parent. Parents are allowed to be present, as long as they are informed on how not to interfere. Siblings should be kept out of the testing room.	Be aware that children will have unexpected needs or wants that will shorten the useful parts of your session. Warn the children when their time is almost up.	comfortable and there should be an activity provided. Waiting times should be minimized if possible.	The moderator should not give opinions. If the moderator feels the child is giving answers in order to please the moderator, the moderator can challenge them in a friendly way. Continually check participants' mood state, noting in the moderator's note for assessing later.
After Testing	Read a child's body language about the product	Follow-up questions should be short and only a few of them.	Children should be thanked in person.	Keep questions neutral, simple, clear, and concrete.

	Older children are able to give reliable ratings on the product. Emphasize how helpful the children were to finding what needs fixed on the product. If able to, reward the children with a token of your appreciation.	When possible, information contributed by participants should be shared with them. Translate children's writing but be sure to keep the original date. Cross check a respondent's answers with their actions.	Avoid using metaphors in case of misinterpretation. Ask for clarification if the response does not seem clear or correct.

3.2 Factors to consider when testing

Even though the guidelines in Table 2 are organized by stage of testing, they can also be divided into different attributes of the test that can be controlled and changed by the researcher or tester. It is important for a researcher to know all the variables that can affect the testing and to know which ones they can manage. These attributes can be categorized into the who, what, when, where, and how of the test, as seen in Table 3.

Table 3. Evolution of Guidelines

Phase of Research	Set-Up and Planning					Introductions					During Testing					After Testing				
	Who	What	When	Where	How	Who	What	When	Where	How	Who	What	When	Where	How	Who	What	When	Where	How
Hanna et al. (1999)	✓	✓	✓	✓						✓	✓	✓	✓		✓	✓	✓			✓
Barendregt & Bekker (2003)	✓	✓								✓		✓			✓		✓			
Markopoulos et al. (2008)	✓	✓	✓	✓	✓					✓		✓	✓	✓	✓					✓
Snitker (2021)			✓	✓	✓					✓			✓		✓					✓

The ‘who’ encompasses everyone involved in testing: children, responsible adults (parents, teachers), those in the testing space, and the moderator. The researcher gets the choice of who is being tested, ideally children that are in the intended age group of the product. The researcher can control how many children are in the testing space at one time. They may want to change this in order to facilitate more feedback (Khanum & Trivedi, 2013) or to test the product in group functions (Coelho & Fernandes, 2013). Parents may not feel comfortable having their children be by themselves with the moderator. If this is the case, parents should be allowed in the testing space if they are trained on how to remain neutral in the testing space (Hanna et al. 1999). The way the moderator acts, or even how they dress can have a big influence on the results of the test. When a researcher wears informal clothes, they will be more likely to develop a relaxed relationship with the participants than when wearing formal clothes in which the child could attribute to authority figures and become more reserved (Druin et al., 1998). The moderator can be a hired usability specialist, the designer, a researcher, or even children’s teachers. It is necessary that whoever is the moderator is trained and heeds the advice as well as possible (Pardo et al., 2006).

The ‘what’ in testing refers to the products and tasks being tested. An instance in which a factor relating to the product in testing can be changed is the order in which products are tested. Serial-position effect is the tendency of a person to remember the first and last stimulus clearer than those in-between. To correct for this effect, the researcher can change the sequence of the products for each participant (Snitker, 2021).

When a test takes place has a more significant influence than one may realize. According to Snitker (2021), it is important to mimic the time in which the product would be used. If the

product is supposed to be used with adult supervision, then that time would be after work hours or weekends. On the other hand, it may be hard to find enough volunteers willing to give up their free hours, then it may be necessary to do the test during school hours in collaboration with a teacher (Markopoulos et al. 2008).

Where usability testing occurs is also important to consider, such as in the field or in a lab. Testing in the field means testing in the actual place, in the physical and social context in which the user is likely to encounter the product. For children, these places tend to be their home or school (Markopoulos et al. 2008). A lab is a space set up by practitioners in which a moderator can observe. Labs vary in complexity, from rooms with one-way mirrors to simple office spaces that have been temporarily transformed (Lewis, 2006). Testing in the field is a superior choice to testing in a lab. The field provides researchers with a more accurate representation of true use than a lab and a higher level of comfortability. The choice between lab and field leads to more choices on the specific surroundings. In a lab, the space should be inviting but not distracting with furniture arrangements that aid the children in focusing on the tasks (Hanna et al. 1999). In the field, challenges are to be expected. One challenge is managing equipment necessary for recording or for display purposes (Warren et al., 2011). When something goes wrong when testing with children, it can prove fatal to the results of the test, data could be lost or the design aspect needing to be tested does not get tested. The researcher should design the study to make sure there are back-ups of all products and technology and, when relying on technology provided by facilities in the field, someone should go there before the test to ensure it works (Markopoulos, Read, et al., 2008). Also, when testing in a formal lab, who and what goes on behind one-way mirrors needs to be explained very clearly in order to be sure the child is aware and all is ethical.

The 'how' embodies the actions and approach during testing. This includes the method(s) chosen for conducting the usability test. The method will indicate what steps need to be taken and what kind of data will be collected. The guidelines provided in this paper do not include the specific best practices for each individual testing method. The evaluative research methods will be further discussed in Section 4.1 Types of Methods. Examples of 'hows' that would be referenced in the guideline are how the moderator introduces themselves, how the children are being asked to do tasks, or how the moderator responds to the children. One example of explaining confidentiality agreements to children is to say the test is "top secret," in that the moderator is asking the child to not share the product information with other people after testing. The children will recognize this comparison and find it fun to act in the role of a "secret agent" (Hanna et al., 1997). This suggestion may be more beneficial after testing to not excite the children and add unwanted positive bias during testing.

4. Comparison of evaluative research methods

Evaluative research methods are specific activities and techniques used to measure and evaluate a product, and as related to Section 3.2 it is part of the 'how' of Table 3. Usability testing is a subset of evaluative research, in that usability testing refers to evaluating a

product or service by testing it with representative users. All products should go through some process of usability testing before being manufactured. The goal of this test is to determine if the product is, "[capable of being] used by humans easily and effectively" (Hornbæk, 2006). There are many methods for usability testing, and 13 of the most prevalent methods found in the literature for testing with children are included in Table 4. These methods can generally be divided into these categories: observation, verbalization, survey, or longitudinal (Markopoulos et al. 2008). In Table 4, data from the literature was compiled in order to compare the individual methods on which ages these tests have been completed, if the moderator needs to be present in the testing room, and the location of the testing, specifically whether it takes place in the field or lab. These factors were chosen because they display the robustness of the method and the similarity to actual use of the product.

Table 4. Comparison of Evaluative Research Methods: Age Groups, Location, Moderator Interaction

	Ages 2-4* (Preconceptual Thought)	Ages 4-7* (Intuitive Thought)	Ages 7-11* (Concrete Operations)	Ages 11-14* (Formal Operations)	Moderator Interacts While Testing**	Field **	Lab**
Method							
Passive Observation		✔ (Biddiss et al., 2013) (Bekker, 2003) (Donker & Reitsma, 2004)	✔ (Sim et al., 2006) (Biddiss et al., 2013)	✔ (Biddiss et al., 2013)			✔
Concurrent Think-Aloud		✔ (Barendregt et al., 2006) (Bekker, 2003) (Donker & Reitsma, 2004) (Barendregt et al., 2007)	✔ (Baauw & Markopoulos, 2004) (Donker & Markopoulos, 2002) (Khanum & Trivedi, n.d.)	✔ (Donker & Markopoulos, 2002) (Als et al., 2005a) (Khanum & Trivedi, n.d.) (Warren et al., 2011)			✔
Retrospective Think-Aloud		✔ (Bekker, 2003)	✔ (Al-Wabil et al., 2010)	✔ (Al-Wabil et al., 2010)			✔
Active Intervention		✔ (Edwards & Benedyk, 2007) (Bekker, 2003)	✔ (Edwards & Benedyk, 2007) (Ognjanovic & Ralls, 2013) (Pardo et al., 2006)		✔		✔
Robotic Intervention		✔ (Markopoulos, Verschoor, et al., 2008)	✔ (Markopoulos, Verschoor, et al., 2008)		✔		✔

Post-Task interview		✔ (Vermeeren et al., 2007)	✔ (Baauw & Markopoulos, 2004) (Vermeeren et al., 2007)	✔ (Donker & Markopoulos, 2002)	✔		✔
Codiscovery		✔ (Bekker, 2003)	✔ (Khanum & Trivedi, n.d.)	✔ (Als et al., 2005a) (Khanum & Trivedi, n.d.) (Als et al., 2005b)		✔	✔
Peer-Tutoring		✔ (Marco et al., 2012) (Edwards & Benedyk, 2007) (Bekker, 2003)	✔ (Edwards & Benedyk, 2007) (Ognjanovic & Ralls, 2013) (Kantosalo & Riihiaho, 2019)			✔	✔
Focus Groups		✔ (Biddiss et al., 2013)	✔ (Rounding et al., 2013) (Biddiss et al., 2013)	✔ (Rounding et al., 2013) (Biddiss et al., 2013)	✔	✔	✔
Questionnaires	✔ (Markopoulos, Read, et al., 2008)	✔ (Barendregt et al., 2006) (Baran, 2009) (J. C. Read & MacFarlane, 2006)	✔ (Sim & Horton, 2012)(Donker & Markopoulos, 2002) (Putnam et al., 2020) (Sim et al., 2006) (Kantosalo & Riihiaho, 2019) (J. C. Read & MacFarlane, 2006)	✔ (Donker & Markopoulos, 2002)	✔	✔	✔
Interview		✔ (Bernhaupt et al., 2007)	✔ (Donker & Markopoulos, 2002) (Bernhaupt et al., 2007) (Kantosalo & Riihiaho, 2019)	✔ (Warren et al., 2011)		✔	✔
Diaries			✔ (Colombo & Landoni, 2014)	✔ (Colombo & Landoni, 2014)		✔	
Wizard-of-Oz		✔ (Marco et al., 2012)	✔ (Höysniemi et al., 2004)		✔		✔

*Ages based on Jean Piaget's stages of development, broken into substages.

**Data gathered for the last three columns (Moderator Interacts while Testing, Field, and Lab) was solely from Markopoulos, Read, et al. 2008 while the previous columns had multiple sources.

Observational methods include participant, passive, and naturalistic observation. Aspects of observational methods are observing and recording the participants' behaviors during the test as well as analyzing the video recording of the session post-test. Participant observation is when the moderator chooses to engage with the participant. Passive participation is the opposite, moderators do not interact or interact minimally with the participants. Naturalistic observation is observation in the field (Markopoulos, Read, et al., 2008). Adaptations can be made to this process. For example, along with gathering data through a video recording, Masood & Thigambaram (2015) tracked eye movements to observe how the children interact with the interface.

Verbalization methods involve keeping a record of things participants say, either spontaneous or encouraged (Markopoulos et al. 2008). When testing with children the verbalization methods that are used are concurrent think-aloud, retrospective think-aloud, active intervention, robotic intervention, post-task interview, codiscovery, and peer tutoring. When using think-aloud, the moderator asks the participants to say their thought-process aloud (Donker & Markopoulos, 2002). Concurrent means during the test while retrospective is afterwards, usually with the aid of video or record of some sort. In active and robotic intervention, the moderator interacts with the tester, asking questions, giving directions, and asking for clarity on opinions, either in person or through a robotic interface (the robot is often just a toy with a speaker in which the moderator can talk to the children from a separate room) (Markopoulos, Read, et al., 2008). Post-task interview is asking questions after each task. They are very practical because they allow immediate observation and verbalization data to be collected (Baauw & Markopoulos, 2004). Co-discovery and peer tutoring involve collaboration between the participants. In co-discovery, multiple testers are evaluating the product at one time (Downey, 2007). For peer-tutoring, one or more participants teach another participant on how to use the product. Peer tutoring is the only testing method created specifically for children. (Markopoulos & Bekker, 2003). The benefit of peer-tutoring is having a role for children to play encourages easy and lively communication because there is not a knowledge or language gap between children as much as there is between a child and an adult (Ognjanovic & Ralls, 2013).

Another verbal evaluative testing method that can only be used for interactive technology products is the Wizard-of-Oz method. This is when a moderator controls all the interactivity on the child's computer through their own computer. This method is necessary for the early stages of development before the technology has been made functional (Markopoulos, Read, et al., 2008). Gesture and movement patterns to computer games can be found through this method, helping designers discover what is most reflective of a child's natural tendencies (Höysniemi et al., 2004).

Survey methods are the processes of asking questions, either through questionnaires or in an interview. Questionnaires are usually given to larger groups of people at one time while interviews are given one at a time (Markopoulos, Read, et al., 2008). Questionnaires are efficient when time is short for feedback; it is possible that interviews can be conducted later to provide clarification when answers on the questionnaires are unclear. Surveys can be modified for children. The Fun Toolkit includes the Smileyometer, the Fun Sorter, and the Again Again Table, all designed to be fun and straightforward (Read, 2008). These tools help involve children and reduce satisficing (Read, 2008), participants giving little thought to answers in order to finish them quickly (Vanette & Krosnick, 2014).

Longitudinal methods, such as the diaries method, take place in a location where there is no evaluator present, usually over many days. Children are asked to document their experiences and opinions in a journal to determine the use of the product in context. Parents can be involved in this method by taking the role of the evaluator, observing and documenting

the child's interaction with the product (Markopoulos, Read, et al., 2008). Longitudinal studies can be used to see whether a product goal holds up through time. In a longitudinal test of PlayCubes TUI technology, researchers concluded the technology works for evaluating children's constructional ability, a foundational goal of the product (Jacoby, 2011). Due to project deadlines, it is likely longitudinal methods are used less frequently in industry.

5. Discussion

Three interesting topics emerged from the synthesis of the literature review that will be discussed next. First, the differences between longitudinal vs. cross-sectional studies. Second, the abundance of digital products compared to analog products in the usability testing with children literature. Third, the examination of the lack of testing with two- to four-year-old children. These areas highlight some opportunities for future research.

5.1 Longitudinal vs. cross-sectional studies

There is some discontinuity in the literature on the context of usability testing. Comparison studies of methods have mostly focused on cross-sectional testing methods, which are multiple participants being tested simultaneously, usually for a short duration (Teti, 2008). Some examples of cross-sectional testing methods include think-aloud, post-task interviews, and questionnaires. In the toy industry, designers are on a very tight schedule of getting the product to market (Johnson, 2001). This could be the reason for the skewed focus on tests that can be completed within a day. What is missing from cross-sectional testing methods is repeated exposure with the product (Markopoulos, Read, et al., 2008). A diary method was used to determine whether enhanced eBooks provided a better learning experience than during leisure reading (Colombo & Landoni, 2014). This exemplifies how context and exposure can be necessary for testing specific aspects of a product. Researchers at Microsoft conducted longitudinal studies during early stages of prototyping. Children were asked to come in for repeated visits over a span of two to three weeks, which equals the approximate amount of time children would use the product over two to three months (Muller & Czerwinski, 1999). This modification helps fit longitudinal studies into the context of actual industry practice. However, the context of actual use was taken out of the study. Longitudinal methods should be explored to determine whether it can be a valid choice for keeping the benefits of exposure and context while fitting within the timeframe of production.

5.2 Analog vs. digital

Products for children can be broadly divided into analog products, digital products, or a combination of analog and digital. Hiniker et al. (2018) uses the terms 'traditional' and 'physical' to describe analog toys: blocks, dolls, craft supplies. Digital refers to any good that can be "digitized or converted into a binary format" (Hui & Chau, 2002). An example of a digital product is a computer game or a mobile phone app. Studying children as users emerged alongside a rising complexity of technology, especially for digital products. Because of this, there is an increase in usability studies and literature for digital products compared to

analog products. Digital products contain complex aspects, including the computer and user interface. Currently, there is a lack of usability testing guidelines for analog products. Even with the decreased technical complexity, analog products have important physical elements that still need to be tested, especially when function and safety are a top priority. Less literature on this topic may be because testing analog products seems more straightforward or more based on 'common sense' than digital products. It could also be because human-computer interaction (HCI) is a more modern topic of discussion, garnering more interest. Whatever the reason, in consideration of children as the user, it could be helpful to study the best methods for testing products from a spectrum of digital to analog. Table 5 categorizes the literature according to what product type the paper studies, either completely digital, digital and analog, or completely analog.

Table 5. Categorization of Digitality of the Products being used in the Literature Under Review

Completely Digital	Digital and Analog	Completely Analog
(Benford et al., 2000) (Sim & Horton, 2012) (Baauw & Markopoulos, 2004) (Donker & Markopoulos, 2002) (Edwards & Benedyk, 2007) (Colombo & Landoni, 2014) (Putnam et al., 2020) (Sim et al., 2006) (Al-Wabil et al., 2010) (Druin et al., n.d.) (Als et al., 2005a) (Ognjanovic & Ralls, 2013) (Barendregt et al., 2003) (J. Read et al., 2009) (Bernhaupt et al., 2007) (Rounding et al., 2013) (Ellis et al., 2008) (Khanum & Trivedi, n.d.) (Als et al., 2005b) (Alkhawajah, 2018) (Hanna et al., 1997) (Barendregt & Bekker, 2003) (Barendregt et al., 2006) (Druin et al., 1997) (Barendregt et al., 2007) (Pardo et al., 2006) (Markopoulos, Verschoor, et al., 2008) (Baran, 2009) (Biddiss et al., 2013) (Druin, 2002) (Buckleitner, 1999) (Masood & Thigambaram, 2015) (Warren et al., 2011) (Alsumait & Al-Osaimi, 2009) (Kantosalo & Riihiaho, 2019) (Donker & Reitsma, 2004) (Hadj-karim-kharrazi et al., 2005) (J. C. Read & MacFarlane, 2006) (J. C. Read, 2008) (Höysniemi et al., 2004)	(Bekker, 2003) (Luo et al., 2018) (Vonach et al., 2016) (Vermeeren et al., 2007) (de Albuquerque & Kelner, 2020) (Jacoby, 2011) (Marco et al., 2012)	(Santos et al., 2019)

5.3 Two- to four-year-old children in user testing

As seen in Table 4, there is a large lack of literature on usability testing with two- to four-year-old children. There are considerable difficulties with this age group, from their language skills to their easy distractibility. However, this age group should still be considered as important as any other testing group. It is important to see how a child naturally interacts with a product. For two- to four-year-old children, they do not have the same mental models that older children or adults hold. Because of this, they are more likely to be creative and use products in a completely unexpected way. These insights provided by the children will help the designer know how to better enable the natural use of the product. A designer could disregard a young child's opinion because the child at that age does not buy the product. However, by only getting adults' opinions, the designer would be doing their product a disservice. Parents and adults want products that their child will remain interested in. To keep the

child interested, the product needs to be easy to use, fun, and safe. This reiterates the importance of observing the child using the prototype, so that the researcher/designer will be able to learn what aspects of the prototype the child really connects with, what aspects they do not, and what factors need to be reconsidered to make it easier for the children to use.

6. Conclusion

"It is important to remember that children are not really 'testing' our prototypes; they are in fact playing, and they will only do so for fun" (Marco et al., 2012). Children are very different from adults, and these differences impact the usability testing for products. It is known that children can become bored, distracted, or act out during testing, but that they are also extremely honest with their feedback (Druin, 2002). When usability testing, honesty is key to creating a better product. Even though children tend to be honest, bias can still occur. Guidelines have been created to promote a better testing experience (as shown in Table 2), ideally with less bias occurring by following these practices. There are many methods of usability testing with children (as shown in Table 3). The researcher must decide which method will generate the best results based on their own needs of data and time available. In the field of usability testing with children, more research can be done regarding which method is most like the context of use of the product and can be conducted in a reasonable timeframe. Also, exploration involving testing and comparing the best practices for products that are digital and analog may provide helpful insight to researchers and designers. As adults, it is important for researchers to be cognizant of the difference between children and adults and to remember that children are their own people with their own opinions, likes, and dislikes (Druin et al., 1997). "In much of our field research we saw that children are natural born artists and writers, architects and philosophers. They are sculptors and poets, dancers, and musicians. Children are not waiting to become these in the future; they are all of those things right now" (Druin et al., 1998). Accepting children for who they are and trusting their instincts to test usability of products in accordance with best practice will give the design team the opportunity to improve their products and learn the unexpected.

7. References

Alkhawajah, A. (2018). Guidelines for Remote Usability Testing of Children's Interactive Products [Ohio University]. https://etd.ohiolink.edu/apexprod/rws_olink/r/1501/10?clear=10&p10_accession_num=ohiou1525711014013404

Als, B. S., Jensen, J. J., & Skov, M. B. (2005a). Comparison of think-aloud and constructive interaction in usability testing with children. Proceeding of the 2005 Conference on Interaction Design and Children - IDC '05, 9–16. https://doi.org/10.1145/1109540.1109542

Als, B. S., Jensen, J. J., & Skov, M. B. (2005b). Exploring Verbalization and Collaboration of Constructive Interaction with Children. In M. F. Costabile & F. Paternò (Eds.), Human-Computer Interaction—INTERACT 2005 (pp. 443–456). Springer. https://doi.org/10.1007/11555261_37

Alsumait, A., & Al-Osaimi, A. (2009). Usability heuristics evaluation for child e-learning applications. 6.

Al-Wabil, A., Al-Husian, L., Al-Murshad, R., & Al-Nafjan, A. (2010). Applying the retrospective think-aloud protocol in usability evaluations with children: Seeing through children's eyes. 2010 International Conference on User Science and Engineering (i-USEr), 98–103. https://doi.org/10.1109/IUSER.2010.5716731

Baauw, E., & Markopoulos, P. (2004). A comparison of think-aloud and post-task interview for usability testing with children. https://doi.org/10.1145/1017833.1017848

Ball, Jonathan. (2019). The Double Diamond: A universally accepted depiction of the design process. https://www.designcouncil.org.uk/news-opinion/double-diamond-universally-accepted-depiction-design-process

Baran, E. (2009). Testing the OLPC (One Laptop per child) XO-1 Interface with Children: A Usability Case Study. 7.

Barendregt, W., & Bekker, M. M. (2003). Guidelines for user testing with children. 4.

Barendregt, W., Bekker, M. M., Bouwhuis, D. G., & Baauw, E. (2006). Identifying usability and fun problems in a computer game during first use and after some practice. International Journal of Human-Computer Studies, 64(9), 830–846. https://doi.org/10.1016/j.ijhcs.2006.03.004

Barendregt, W., Bekker, MathildeM., Bouwhuis, DonG., & Baauw, E. (2007). Predicting effectiveness of children participants in user testing based on personality characteristics. Behaviour & Information Technology, 26(2), 133–147. https://doi.org/10.1080/01449290500330372

Barendregt, W., Bekker, T., & Speerstra, M. (2003). Empirical evaluation of usability and fun in computer games for children.

Bekker, M. M. (2003). Assessing Usability Evaluation Methods On Their Effectiveness to Elicit Verbal Comments From Children Subjects. 9.IDC '03: Proceedings of the 2003 conference on Interaction design and children. July 2003. Pages 41–49. https://doi.org/10.1145/953536.953544

Benford, S., Bederson, B. B., Åkesson, K.-P., Bayon, V., Druin, A., Hansson, P., Hourcade, J. P., Ingram, R., Neale, H., O'Malley, C., Simsarian, K. T., Stanton, D., Sundblad, Y., & Taxén, G. (2000). Designing storytelling technologies to encouraging collaboration between young children. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 556–563. https://doi.org/10.1145/332040.332502

Bernhaupt, R., Schwaiger, D., Riegler, S., & Enthaler, D. (2007). Evaluating children's gaming experiences. Proceedings of the International Conference on Advances in Computer Entertainment Technology - ACE '07, 224. https://doi.org/10.1145/1255047.1255096

Biddiss, E., McPherson, A., Shea, G., & McKeever, P. (2013). The Design and Testing of Interactive Hospital Spaces to Meet the Needs of Waiting Children. HERD: Health Environments Research & Design Journal, 6(3), 49–68. https://doi.org/10.1177/193758671300600305

Buckleitner, W. (1999). The State of Children's Software Evaluation—Yesterday, Today, and in the 21st Century. Information Technology in Childhood Education Annual, 1999(1), 211–220.

Coelho, D., & Fernandes, S. (2013). Toy Design—A Methodological Perspective. THE INTERNATIONAL JOURNAL OF DESIGNED OBJECTS, 7.

Colombo, L., & Landoni, M. (2014). A diary study of children's user experience with EBooks using flow theory as framework. Proceedings of the 2014 Conference on Interaction Design and Children, 135–144. https://doi.org/10.1145/2593968.2593978

de Albuquerque, A. P., & Kelner, J. (2020). Human-Centered Design Tools for Smart Toys (P. C. K. Hung, Ed.; Vol. 11894). Springer International Publishing. https://doi.org/10.1007/978-3-030-38651-1

Dependents 2 | Internal Revenue Service. (2021). Retrieved November 15, 2021, from https://www.irs.gov/faqs/filing-requirements-status-dependents/dependents/dependents-2

Donker, A., & Markopoulos, P. (2002). A comparison of think-aloud, questionnaires and interviews for testing usability with children. In, 305–316.

Donker, A., & Reitsma, P. (2004). Usability testing with young children. Proceeding of the 2004 Conference on Interaction Design and Children Building a Community - IDC '04, 43–48. https://doi.org/10.1145/1017833.1017839

Downey, L. L. (2007). Group Usability Testing: Evolution in Usability Techniques. Journal of Usability Studies, 133–144.

Druin, A. (2002). The Role of Children in the Design of New Technology. Behaviour and Information Technology, 38.

Druin, A., Bederson, B., Boltman, A., Miura, A., Knotts-Callahan, D., & Platt, M. (1998). Chapter 3: Children as Our Technology Design Partners+. 9.

Druin, A., Stewart, J., Proft, D., Bederson, B., & Hollan, J. (1997). KidPad: A design collaboration between children, technologists, and educators. Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, 463–470. https://doi.org/10.1145/258549.258866

Dumas, J. S., & Fox, J. E. (2007). Usability Testing. 22.

Edwards, H., & Benedyk, R. (2007). A comparison of usability evaluation methods for child participants in a school setting. Proceedings of the 6th International Conference on Interaction Design and Children - IDC '07, 9. https://doi.org/10.1145/1297277.1297280

Ellis, K., Quigley, M., & Power, M. (2008). Experiences in Ethical Usability Testing with Children: Journal of Information Technology Research, 1(3), 1–13. https://doi.org/10.4018/jitr.2008070101

Gustafsson, D. (2019). Analysing the Double diamond design process through research & implementation. https://aaltodoc.aalto.fi:443/handle/123456789/39285

Hadj-karim-kharrazi, H., B, I., & Blustein, P. D. J. (2005). Usable Guidelines for Usability Testing with Children Student.

Hanna, L., Risden, K., & Alexander, K. (1997). Guidelines for usability testing with children. Interactions, 4(5), 9–14. https://doi.org/10.1145/264044.264045

Hiniker, A., Lee, B., Kientz, J. A., & Radesky, J. S. (2018). Let's Play! Digital and Analog Play between Preschoolers and Parents. Association for Computing Machinery. https://doi.org/10.1145/3173574.3174233

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. International Journal of Human-Computer Studies, 64(2), 79–102. https://doi.org/10.1016/j.ijhcs.2005.06.002

Höysniemi, J., Hämäläinen, P., & Turkki, L. (2004). Wizard of Oz prototyping of computer vision based action games for children. Proceeding of the 2004 Conference on Interaction Design and Children Building a Community - IDC '04, 27–34. https://doi.org/10.1145/1017833.1017837

Hui, K. L., & Chau, P. Y. K. (2002) Classifying digital products. Communications of the ACM. https://dl.acm.org/doi/10.1145/508448.508451

Jacoby, S. (2011). Playcubesa Longitudinal Study of Tangible User Interface for Assessment and Intervention of Dynamic Constructional Processes Among Typically Developing Children [Ph.D., University of Haifa (Israel)]. https://www.proquest.com/docview/2586971497/abstract/9A6CAD5A71C94D4CPQ/1

Jm3. (2016). English: Wikipedia's complete (as of 2016) list of cognitive biases, arranged and designed by John Manoogian III (jm3). Categories and descriptions originally by Buster Benson. Own work, categorization by Buster Benson[1]. https://commons.wikimedia.org/wiki/File:The_Cognitive_Bias_Codex_-_180%2B_biases,_designed_by_John_Manoogian_III_(jm3).png

Johnson, M. E. (2001). Learning from Toys: Lessons in Managing Supply Chain Risk from the Toy Industry. California Management Review, 43(3), 106–124. https://doi.org/10.2307/41166091

Kantosalo, A., & Riihiaho, S. (2019). Usability Testing and Feedback Collection in a School Context: Case Poetry Machine. Ergonomics in Design, 27(3), 17–23. https://doi.org/10.1177/1064804618787382

Khanum, M. A., & Trivedi, M. C. (2013). Exploring Verbalization and Collaboration during Usability Evaluation with Children in Context. 7.

Kienzler, M. (2018). Value-based pricing and cognitive biases: An overview for business markets. Industrial Marketing Management, 68, 86–94. https://doi.org/10.1016/j.indmarman.2017.09.028

Kinsey, M. J., Gwynne, S. M. V., Kuligowski, E. D., & Kinateder, M. (2019). Cognitive Biases Within Decision Making During Fire Evacuations. Fire Technology, 55(2), 465–485. https://doi.org/10.1007/s10694-018-0708-0

Kuniavsky, M. (2003). Observing the User Experience: A Practitioner's Guide to User Research. Elsevier.

Lauff, C. A., Kotys-Schwartz, D., & Rentschler, M. E. (2018). What is a Prototype? What are the Roles of Prototypes in Companies? Journal of Mechanical Design, 140(6), 061102. https://doi.org/10.1115/1.4039340

Luo, S., Wang, Y., Xiong, N., Shan, P., & Zhou, Y. (2018). An Interactive Smart Music Toy Design for Children. In N. Streitz & S. Konomi (Eds.), Distributed, Ambient and Pervasive Interactions: Understanding Humans (pp. 372–390). Springer International Publishing. https://doi.org/10.1007/978-3-319-91125-0_31

Marco, J., Cerezo, E., & Baldassarri, S. (2012). Tangible interaction and tabletops: New horizons for children's games. International Journal of Arts and Technology, 5(2/3/4), 151. https://doi.org/10.1504/IJART.2012.046272

Markopoulos, P., & Bekker, M. (2003). Interaction design and children. Interacting with Computers, 15(2), 141–149. https://doi.org/10.1016/S0953-5438(03)00004-3

Markopoulos, P., Read, J. C., MacFarlane, S., & Hoysniemi, J. (2008). Evaluating Children's Interactive Products: Principles and Practices for Interaction Designers. Elsevier.

Markopoulos, P., Verschoor, Y., Ijsselsteijn, W., & de Ruyter, B. (2008). Testing interactive products with the robot intervention method. Proceedings of the 7th International Conference on Interaction Design and Children - IDC '08, 117. https://doi.org/10.1145/1463689.1463735

Masood, M., & Thigambaram, M. (2015). The Usability of Mobile Applications for Pre-schoolers. Procedia - Social and Behavioral Sciences, 197, 1818–1826. https://doi.org/10.1016/j.sbspro.2015.07.241

Muller, M., & Czerwinski, M. P. (1999). Organizing usability work to fit the full product range. Communications of the ACM, 42(5), 87–90. https://doi.org/10.1145/301353.301436

Natesan, D., Walker, M., & Clark, S. (2016). Cognitive Bias in Usability Testing. Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care, 5(1), 86–88. https://doi.org/10.1177/2327857916051015

Ognjanovic, S., & Ralls, J. (2013). Don't talk to strangers!: Peer tutoring versus active intervention methodologies in interviewing children. CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13, 2337. https://doi.org/10.1145/2468356.2468761

Pardo, S., Vetere, F., & Howard, S. (2006). Teachers' involvement in usability testing with children. Proceeding of the 2006 Conference on Interaction Design and Children - IDC '06, 89. https://doi.org/10.1145/1139073.1139097

Putnam, C., Puthenmadom, M., Cuerdo, M. A., Wang, W., & Paul, N. (2020). Adaptation of the System Usability Scale for User Testing with Children. Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, 1–7. https://doi.org/10.1145/3334480.3382840

Read, J. C. (2008). Validating the Fun Toolkit: An instrument for measuring children's opinions of technology. Cognition, Technology & Work, 10(2), 119–128. https://doi.org/10.1007/s10111-007-0069-9

Read, J. C., & MacFarlane, S. (2006). Using the fun toolkit and other survey methods to gather opinions in child computer interaction. Proceeding of the 2006 Conference on Interaction Design and Children - IDC '06, 81. https://doi.org/10.1145/1139073.1139096

Read, J., MacFarlane, S., & Casey, C. (2009). Endurability, Engagement and Expectations: Measuring Children's Fun. Interaction Design and Children.

Rounding, K., Tee, K., Wu, X., Guo, C., & Tse, E. (2013). Evaluating interfaces with children. Personal and Ubiquitous Computing, 17(8), 1663–1666. https://doi.org/10.1007/s00779-012-0521-6

Sanders, E. (2002). From user-centered to participatory design approaches. In Design and the Social Sciences: Making Connections. (pp. 1–7). https://doi.org/10.1201/9780203301302.ch1

Santos, A. D. P. dos, Junior, G. B., & Medola, F. O. (2019). Innovation in the Design of Inclusive Toys: Development and Evaluation of a Prototype for Visually Impaired Children. Strategic Design Research Journal, 12(3), 338–360.

Schwenk, C. H. (1986). Information, Cognitive Biases, and Commitment to a Course of Action. Academy of Management Review, 11(2), 298–310. https://doi.org/10.5465/amr.1986.4283106

Sim, G., & Horton, M. (2012). Investigating children's opinions of games: Fun Toolkit vs. This or That. Proceedings of the 11th International Conference on Interaction Design and Children, 70–77. https://doi.org/10.1145/2307096.2307105

Sim, G., MacFarlane, S., & Read, J. (2006). All work and no play: Measuring fun, usability, and learning in software for children. Computers & Education, 46(3), 235–248. https://doi.org/10.1016/j.compedu.2005.11.021

Snitker, T. V. (2021). User Research with Kids: How to Effectively Conduct Research with Participants Aged 3-16. Apress. https://doi.org/10.1007/978-1-4842-7071-4

Teti, D. M. (2008). Handbook of Research Methods in Developmental Science. John Wiley & Sons.

Usabilitytesting-ral.pdf. (n.d.). Retrieved November 2, 2021, from http://sistemas-humano-computacionais.wdfiles.com/local--files/capitulo:modelagem-e-simulacao-de-sistemas-humano-computacio/usabilitytesting-ral.pdf

Vermeeren, A., Bekker, M. M., Kesteren, I. E. H. van, & Ridder, H. D. (2007, September 1). Experiences with Structured Interviewing of Children During Usability Tests. Proceedings of HCI 2007 The 21st British HCI Group Annual Conference University of Lancaster, UK. https://doi.org/10.14236/ewic/HCI2007.14

Vonach, E., Ternek, M., Gerstweiler, G., & Kaufmann, H. (2016). Design of a Health Monitoring Toy for Children. Proceedings of the The 15th International Conference on Interaction Design and Children, 58–67. https://doi.org/10.1145/2930674.2930694

Warren, S., Jones, G., & Lin, L. (2011). Usability and Play Testing: The Often Missed Assessment. In L. Annetta & S. C. Bronack (Eds.), Serious Educational Game Assessment (pp. 131–146). SensePublishers. https://doi.org/10.1007/978-94-6091-329-7_8

About the Authors:

Andria Banker is a graduate student in the Product Design Program at the University of Minnesota. Prior to graduate school, she received her Bachelor of Fine Arts from Eastern Kentucky University.

Carlye Lauff is an Assistant Professor of Product Design at the University of Minnesota. Her research focuses on understanding how designers engage in the product development process and then improving tools and methods to support them, such as evaluative research techniques and prototyping strategies.