UNIVERSITÉ DU LUXEMBOURG

PhD-FHSE-2024-013

The Faculty of Humanities, Education and Social Sciences

DISSERTATION

Defence held on 27/03/2024 in Esch-sur-Alzette, Luxembourg
to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG EN PSYCHOLOGIE

by

FLORENCE KRISTIN LEHNERT

Born on 22 December 1985 in Lebach, Germany

MEASURING CHILDREN'S USER EXPERIENCE WITH E-ASSESSMENTS: IMPLICATIONS FOR A BETTER INTERPRETATION OF UX EVALUATION METHODS FOR SCHOOL-AGED CHILDREN

Dissertation defence committee

Dr Pedro Cardoso-Leite, dissertation supervisor
Associate Professor, University of Luxembourg

Dr Gavin Sim
Reader, University of Central Lancashire

Dr Antoine Fischbach, Chair
Associate Professor, University of Luxembourg

Dr Carine Lallemand, Vice-Chair
Research Scientist, University of Luxembourg

Dr Bente Rigmor Walgermo
Associate Professor, University of Stavanger

Measuring Children’s User Experience with E-Assessments:

Implications for a Better Interpretation of UX Evaluation Methods for School-Aged Children

Florence Kristin Lehnert

Abstract

Electronic assessment (e-assessment) is becoming increasingly popular in modern education systems and is emerging as a major and relevant method for evaluating students. Among the variety of new approaches and tools that have been introduced for e-assessment, the absence of well-defined protocols for checking the fairness and reliability of evaluations for every student remains a major problem. To check the credibility of e-assessments, it is paramount to enhance the understanding of children's User Experience (UX), as learners possess limited cognitive resources, and substantial variations exist across different age groups. The interface's usability, characterized by its ease of comprehension, ease of use, and ease of learning, can guarantee a positive UX for children. Children can contribute to the legitimacy of assessment findings only when they are actively permitted to engage in the system and successfully demonstrate their expertise and skills. It is therefore essential to integrate user-centric technology into e-assessments to ensure equitable opportunities for all users.

Thus, the primary objective of this doctoral dissertation is to enhance the methodological understanding of UX evaluation methods for young school-aged children in the context of e-assessments. As a first step, we conducted a systematic review (272 full papers) that allowed us to identify an absence of interaction design techniques that can provide an accurate assessment of the fundamental concepts of UX for students between the ages of 6 and 12 (Research Paper 1). Furthermore, we identified the need for methods that might be employed before or after the evaluation setting to prevent any interference with the children's performance during the assessment. When we began our investigation, we opted to utilize questionnaires as the principal instrument for evaluating children's UX. Nevertheless, it became apparent that the scale—which must be carefully designed when aimed at children, particularly those in our age bracket—was restricted in range and thus required more careful assessment to improve its validity and reliability. Therefore, we decided to dedicate resources to analyzing the scale design as a first action (Research Paper 2) by conducting a cognitive interview study ($N = 25$) and administering a digital large-scale questionnaire in a real assessment scenario ($N = 3{,}263$). Upon scrutinizing the many biases present in this research, we decided to employ experts' perspectives to tackle the issues effectively and provide a valuable contribution to integrating information in a multidisciplinary area (Research Paper 3). We derived 10 heuristics on the basis of an evidence-based corpus of 506 guidelines, the evaluation of several domain experts ($N = 24$), and one heuristic evaluation workshop ($N = 2$). These heuristics may be utilized to assess the UX and usability elements of e-assessments among children aged 6 to 12 years.

Along with our studies, our inquiry found that the methods used to evaluate children's UX needed more careful examination. We advocate for the enhancement and advancement of scales that are both enjoyable and easy for children to use. Furthermore, we suggest that future research efforts incorporate various stakeholder roles for adults and children to promote a paradigm shift by pushing forward participatory design trends. Overall, with this doctoral dissertation, we provide an essential step for improving comprehensiveness and future approaches for evaluating children's UX in an e-assessment context.

Keywords: User Experience, E-assessment, Children, Methods

Acknowledgments

Behind this achievement are powerful mentors, colleagues, and friends who inspired me to move forward and who helped shape my future career path. Thus, I want to thank everyone who crossed my path during this journey.

In particular, I thank Dr. Vincent Koenig for his wise mentorship, expertise, and trust in me. For most of my PhD journey, he was the rock in the waves, and I felt sadness that he had to leave during the last chapter. I thank Prof. Pedro Cardoso-Leite for his open-mindedness and proactive supervision during the previous few months. I express my gratitude to Dr. Carine Lallemand for sharing her profound expertise in User Experience (UX) and for her advice throughout the years. I thank Prof. Antoine Fischbach for his valuable feedback, support, and help in navigating the education jungle.
Moreover, I thank Dr. Gavin Sim for his sound expertise in Child-Computer Interaction (CCI) and education, and for working with me on the last pieces of this work. I also want to thank Dr. Bente Rigmor Walgermo for reading my dissertation and supporting my defense. Additionally, I would like to thank my co-authors, Prof. Panos Markopoulos, Dr. Jasmin Niess, and Sophie Doublet for their guidance, support, and inspiration. I also thank Jane Zagorski for her professional editing skills, which she applied to almost all chapters of this doctoral dissertation.

Moreover, I thank the HCI Team of the University of Luxembourg for their feedback and hearty discussions; in particular, my thanks go to Björn, Chris, Kerstin, Luce, Margault, Verena, and Vincent.
Last and most importantly, I thank my mother, beloved friends and family who supported me and made this journey joyful: Sayang, thank you for having my back and empowering me in so many aspects; my father, thank you for inspiring and strengthening me to do my best possible work and have faith in myself; and my son, you are the driving force, catalyst, and inner voice that drives me to go beyond my abilities and accomplish something great.

Thank you!

Table of Contents

Chapter 1:

General Introduction

1. Introduction

1.1. E-assessments and current trends in education

Assessment has always played a significant role in educational practices, as it helps educators understand and evaluate student learning and identify areas for improvement. At a policy level, it facilitates well-informed decisions about teaching practices and curriculum development. Good performances in high-stakes assessments give students access to further educational opportunities and employment (Ridgway et al., 2004). Important considerations about what students need to know, how much of it can be taught, and how best to promote learning can be found in school assessment programs. Thus, assessment plays four different roles in education: (a) It helps with future learning (formative), (b) it provides an overview of how well students did in a class (summative), (c) it provides a way to select people on the basis of their qualifications (certification), and (d) it enables individuals to determine how well a class or program worked overall (evaluative) (Hornby, 2003). Students often prioritize summative assessments above formative assessments, as summative assessments directly impact students' final grade (Taras, 2002). However, all types of assessments are closely linked to learning (Mackenzie, 2003) and offer alignment between teaching and assessment methods (Ashton & Thomas, 2006). Consequently, evaluation, pedagogy (i.e., teaching), and learning are interconnected. Thus, teaching methods and assessments should be properly aligned with the learning activities as outlined in a course's objectives. We can foresee that only a consistent teaching system supports student learning.

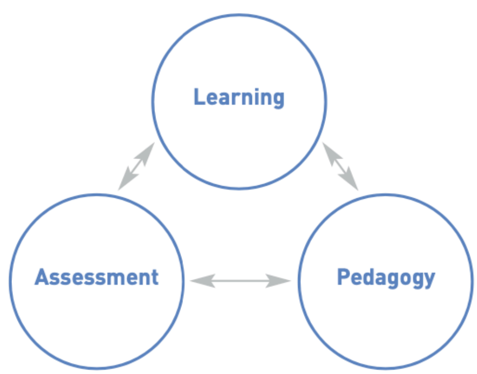


Figure 1. The relationships between assessment, learning, and pedagogy, extracted from (Ridgway et al., 2004), adapted from (Pellegrino et al., 2001).

For many years, summative assessments in particular have usually consisted of timed, closed-book, pen-and-paper examinations. Thus, authors have expressed doubt about whether traditional forms of assessment are correctly aligned with teaching and learning (i.e., a multiple-choice exam to evaluate skills) and have thus suggested that assessments should be concerned with how a learning activity is performed rather than focusing on the outcome of the activity (Ridgway et al., 2004).

"E-assessment must not simply invent new technologies which recycle our current ineffective practices." Martin Ripley, QCA, 2004 (Ridgway et al., 2004) (p. 2).

With the rise of technology, educational practices have begun to incorporate digital tools. Students study with computers (e.g., word processors and graphical calculators), yet they're assessed on class material with pen and paper. To bridge this divide, electronic assessment (e-assessment) is gaining significant traction in current education systems, emerging as a prominent and relevant trend in education and the professional worksphere. As the name suggests, e-assessment involves the use of electronic instruments, specifically computers, instead of traditional paper and pencil methods for evaluation (Jordan, 2013). Other authors (Joint Information Systems Committee, 2007) have provided a narrower definition of e-assessment and have concluded that it is an end-to-end electronic assessment process in which Information Communication Technology (ICT) is used for the entire assessment process, from the presentation of questions to the recording of the learners' responses. The literature is replete with similar terms that refer to the same concept, including technology-enhanced or technology-enabled assessment and computer-assisted, computer-based, or computer-aided assessment (Jordan, 2013; Sim et al., 2007). In this dissertation, I use the term electronic assessment (e-assessment), as it includes any use of a computer as part of any activity related to the assessment (Joint Information Systems Committee, 2007).

In the 1990s, researchers predicted four generations of e-assessment (which they referred to as computer-based evaluation) (Bunderson et al., 1988). Their predictions were entirely accurate but overly optimistic about the timing (Redecker & Johannessen, 2013). First-generation CBA, or computerized testing, focused solely on digitizing paper-and-pencil tests (Bunderson et al., 1988). Additionally, the introduction of the World Wide Web in the 1990s resulted in significant changes in several industries, mainly in education (Llamas-Nistal et al., 2013). From then onwards, many companies implemented e-assessment systems until the Joint Information Systems Committee (JISC) introduced joint principles and advice for e-assessment in England, Wales, and Northern Ireland (Joint Information Systems Committee, 2007). The second generation, computerized adaptive testing, went one step further and modified test difficulty, duration, and material on the basis of student performance (Bunderson et al., 1988). In the third generation of CBA, continuous measurement focused on taking continuous, unobtrusive performance measurements and keeping them in places such as log files (Bunderson et al., 1988). The fourth generation of intelligent measurement focused on centralizing the production of intelligent scoring, interpreting individual profiles, and recommending to learners and teachers through knowledge bases and inferencing methods (Bunderson et al., 1988). For instance, in 2009, Cisco, Intel, and Microsoft transformed the Education by Assessing and Teaching 21st Century Skills approach (Kozma, 2009), in which the learning experience is tailored for each student (Bunderson et al., 1988).

The development and ever-growing use of e-assessments have spread through many different contexts (e.g., educational, professional). Thus, in an educational context, we can synthesize the numerous benefits of e-assessment for two groups of individuals: those engaged in its creation (e.g., educators, administrators, designers, and developers) and those who undergo the assessments (i.e., the students). From the creators, the flexibility in providing electronic feedback can, for example, offer new possibilities for addressing multimodal learning, which has been shown to increase students' self-reflection (Whitelock et al., 2015) and motivation (Rolim & Isaias, 2019; St-Onge et al., 2022). Moreover, e-assessment can be used to monitor students' progress longitudinally and offer early identification of students with learning difficulties (Rolim & Isaias, 2019). Other studies have mentioned more advantages, such as time-saving options (Dermo, 2007), improved data management and analysis, support in the examination processes, and technical enhancement of the quality of tests by improving the reliability of scoring (Jordan, 2013). Alongside increased reliability, e-assessment can offer advantages that are particularly useful for large classes (Whitelock & Brasher, 2006), and it can add value and enable practitioners to make more productive use of their time (JISC & Team, 2010; Jordan, 2013).

From the student's perspective, e-assessment provides novel opportunities for innovation in testing and evaluation (Bennett, 1998; Farrell & Rushby, 2016; Kyllonen, 2009). First and foremost, e-assessments provide the opportunity for rapid or even immediate feedback (Daly et al., 2010; Jordan, 2013; Rolim & Isaias, 2019), enabling students to identify areas they need to develop and improve (especially relevant for formative assessment) (Ridgway et al., 2004; St-Onge et al., 2022). Interactive feedback approaches provided by technology-supported assessment or e-assessment can result in a positive learning experience (Marriott, 2009), increased competency (Gilbert et al., 2011), and enhanced reflective practice among students (Wongvorachan et al., 2022). Additional student benefits include the opportunity to study at one’s own pace regardless of location, providing more flexibility (Osuji, 2012; Williams & Wong, 2009). Several studies have confirmed that students prefer e-assessment over pen-and-paper-based evaluations (Donovan et al., 2007; Llamas-Nistal et al., 2013; Tubaishat et al., 2006).

Consequently, e-assessment can contribute to a more engaging and effective learning environment by changing how learning, teaching, and assessment occur, resulting in a more efficient and effective education system. The construction of innovative item formats (Parshall et al., 2000) (e.g., dynamic or interactive multimedia items) that are not easily assessed by pen-and paper-based evaluations (JISC & Team, 2010) is an advantage of e-assessments.

In addition to the many benefits that e-assessment provides to education, it is necessary to thoroughly analyze the circumstances and reasons for its use and the advantages of a digital system over a typical pen-and-paper assessment. There is empirical evidence that paper-based and computer-based tests have divergent outcomes. The term used to describe these findings is the test mode effect (Clariana & Wallace, 2002). However, the impact of this phenomenon is generally disregarded when paper-based tasks are converted into computer-based tasks, and users' performance or evaluation experience is usually not taken into consideration (Parshall et al., 2000). Weinerth (2015) called attention to concerns about the validity and fairness of e-assessment for all individuals. Weinerth (2015) demonstrated that e-assessment outcomes can be influenced by confounding variables, many of which are directly associated with computer use. Individual variations in computer literacy, which refers to a person’s proficiency in using technology, might lead to distinct outcomes that are unrelated to the problem-solving task (Weinerth, 2015). Instead, these outcomes may be influenced by the unique usability of the software environment as perceived by each individual (Weinerth, 2015). This observation is in line with the conclusions drawn by other researchers (Clariana & Wallace, 2002; Farrell & Rushby, 2016). Thus, in order to mitigate the effects of Human-Computer Interaction (HCI) problems and enhance the accuracy of e-assessments, it is imperative to consider the user experience (UX) of e-assessment users. Moreover, if the interface's usability is exceptional (e.g., if it is easy to comprehend, effortless to use, and simple to obtain), it might ensure a positive UX, allowing the user to focus on the main job of the evaluation (Weinerth, 2015).

Therefore, with this dissertation, my coauthors and I explored the notion of UX and how it corresponds to the challenges of designing e-assessment interfaces.

1.2. The importance of introducing UX in e-assessments

Quality Indicators of Technology have traditionally been measured by usability (Nielsen, 1994). ISO 9241-11 defines usability as the use of a product "by specified users to achieve specific goals with effectiveness, efficiency, and satisfaction in a specified context of use" (International Organization for Standardization [ISO], 2018, definition 3.1). Usability focuses on product design, particularly the user interface (UI) as the "channel for interactions" (Weinerth, 2015, p. 20). It focuses on how aspects of the product design (rather than a person's ability to use a system) improve task performance (Weinerth, 2015). Researchers have argued that usability alone cannot represent the vast spectrum of humans' experiences with technology (Hassenzahl, 2018; McCarthy & Wright, 2004). More recently, the concept of UX was introduced to fill in the incomplete notions of usability.

UX is described as "a user's perceptions and responses that result from the use or anticipated use of a system, product or service" in ISO 9241-11 (International Organization for Standardization [ISO], 2018, definition 3.2.3). UX defines perceptions and reactions as "the users' emotions, beliefs, preferences, perceptions, comfort, behaviors, and accomplishments that occur before, during, and after use" (International Organization for Standardization [ISO], 2018, definition 3.2.3). Hassenzahl and Tractinsky (2016) define UX as: "A consequence of a user's internal state (predispositions, expectations, needs, motivation, mood, etc.), the characteristics of the designed system (e.g., complexity, purpose, usability, functionality, etc.) and the context (or the environment) within which the interaction occurs (e.g., organizational/social setting, meaningfulness of the activity, voluntariness of use, etc.)" (p. 95). The differences between usability and UX have been debated in the past (Lallemand et al., 2015).

Other experts have adopted a broader perspective on the old usability notions in what they call the *third wave of HCI* (Bødker, 2006, 2015) or *third paradigm of HCI* (Harrison et al., 2007). Therefore, the overall UX must be considered a holistic construct (Lallemand & Koenig, 2017) to comprehensively understand an individual's use of technology (Bargas-Avila & Hornbæk, 2011). This UX is typically divided into *pragmatic* and *hedonic aspects* (Hassenzahl, 2001). The pragmatic or *ergonomic* (Hassenzahl, 2001); or *instrumental* (Thüring & Mahlke, 2007) part of UX focuses on efficacy and efficiency in achieving a goal. The pragmatic dimension usually includes perceived usability (Lallemand, 2015). The hedonic component of UX goes beyond utilitarian concerns and covers aesthetics, emotions, and joy in using technology (Hassenzahl & Tractinsky, 2006).

More recently, HCI researchers have begun to focus on yet another defining component of UX, which they call the *eudaimonic* dimension (Mekler & Hornbæk, 2016) and have suggested that eudaimonia should be added to the definition of UX (Mekler & Hornbæk, 2019). The terms eudaimonic and hedonic seem similar at first sight. However, there are nuanced differences between them with respect to need fulfillment. For the hedonic component, users seek well-being via pleasure and joy, whereas in the eudaimonic component, users seek well-being through deeper human aims, such as excellence, purpose, or values (Huta & Ryan, 2010; Huta & Waterman, 2014). The concepts resemble the need-driven theories in psychology (e.g., Maslow's hierarchy, comprising a five-tier model of human needs, often depicted as hierarchical levels within a pyramid) (Maslow, 1958). The hedonic component could be associated with lower and less long-lasting psychological conditions (e.g., pleasure), whereas the eudaimonic element suggests higher esteem-based needs (e.g., self-actualization). Therefore, it might not be surprising that authors have previously linked the eudaimonic component to learning (Mekler & Hornbæk, 2016), as it emphasizes the importance of positive and long-lasting experiences for learning and assessment (Pekrun, 2011).

Investigation of the UX in any learning context is important because learners have limited mental resources (Baddeley & Hitch, 1974; Miller, 1956). This notion stems from Cognitive Load Theory (CLT), which emphasizes that human working memory is limited, and technology should be designed with these limitations in mind (Sweller, 1988). In CLT, three kinds of cognitive load are explained in more detail: intrinsic, extraneous, and germane cognitive load (Sweller et al., 1998, 2019). Germane cognitive load is especially applicable in the context of e-assessment, as it refers to complexity that is beneficial for learning and assessment purposes (Hollender et al., 2010) because learners need to actively integrate new information into existing schemata. However, extraneous cognitive load, which refers to complexity derived from the instructions, task, design, or other aspects that need to be processed by learners (Hollender et al., 2010; Sweller et al., 1998), is also relevant to e-assessment, as it can determine how easy or difficult it is to use a digital product.

The overarching strategy of several capacity models, referred to as *resource models*, is that the required resources for a work, known as task demand, are not static and that any freeing up of cognitive resources might aid performance (e.g., e-assessment performance) (Wickens, 2002). In summary, any optimized tool for addressing the notions of usability and UX should help learners perform their primary task (e.g., the assessment). Several resources mandate usability notions, such as the International

Guidelines for Computer-Based and Internet-Delivered Testing (International Test Commission, 2006). However, the focus is solely on the system's usability, not the previously described UX concepts.

In the following, we outline the challenges involved in investigating children’s UX in particular.

1.3. Investigating children’s UX in an e-assessment context

“Computers for kids need to be fun like a friend, but can make me smart for school. They should also be friendly like my cat. The real thing is that they shouldn’t make me have to type since I don’t like that. I can talk much better!”

-Quote from an 8-year-old child, extracted from Druin (2002)-

According to a recent survey, 42% of youngsters aged 5 to 7 in the United Kingdom have tablets (Burns & Gottschalk, 2019). Similarly, digital tools have been included in educational practices and integrated into e-assessments that are closely aligned with the content of the OECD's latest Programme for International Student Assessment (PISA) report (*About PISA*, n.d.). PISA seeks to evaluate educational systems around the globe by assessing students' knowledge and abilities at the age of 15 (*About PISA*, n.d.). In the 1990s, interaction designers’ and researchers’ recognition of the growing interest and need to involve children as users and participants in research gave rise to the discipline of child-computer interaction (CCI) (Druin, 1999). Developmental psychology, learning science, computer science, and interface design are interdisciplinary fields impacting CCI (Hourcade, 2015). Research done in CCI mainly explores the design of interactive technologies for children while also looking at the impact of these technologies on children and society (Hourcade, 2015). Simultaneously, technology designers are trying to devise innovative approaches for designing and assessing how children engage with emerging technology. In child-centered research, methodologies that have been developed for children have frequently stemmed from experiences with the adult demographic. Hence, scholars have emphasized the need to gain a more comprehensive understanding of how such methodologies need to be adjusted to address a nonadult population (Markopoulos & Bekker, 2003). Therefore, it is crucial to collect further scientific and methodological evidence on how to best evaluate interactive systems designed for children (Lehnert et al., 2022; Markopoulos & Bekker, 2003).

A fair e-assessment system should consider a wide range of user characteristics, including age, culture, and other requirements (Reich & Petter, 2009). Age is an essential factor to consider among these traits because most recent research on UX and the methodologies underlying it have been conducted on adult users (Fitton et al., 2014). Due to the substantial differences that exist between children and adults in cognitive, physical, and emotional characteristics, it is important to carefully consider how to develop methods for assessing children. As an example, preschool-aged children have a diminished capacity for sustained focus and possess cognitive capabilities that are still in the process of maturing (Piaget, 1955). Additionally, preschoolers may have a different degree of technology literacy than adults. Thus, designers need to consider a particular age group's developmental needs, preferences, and capabilities when designing systems to ensure that they will be effective and will be able to engage the group. Examples may encompass the development of intuitive and user-friendly interfaces, the incorporation of visual aids and feedback mechanisms, and the provision of clear and concise instructions and assistance. Therefore, system designers need to consider the unique needs of children, especially when it comes to drawing conclusions from how they engage with a system, such as an e-assessment platform. Only when children are allowed to actively participate in the system can they effectively showcase their knowledge and abilities which is required to ascertain the validity of the assessment outcomes. User-centered technology is critical in e-assessment to provide fair opportunities for all users.

Therefore, to lessen the effect of HCI problems and enhance the validity of electronic assessments, it is vital to observe the UX of children while they use the e-assessment platforms. An interface with high usability, characterized by its ease of comprehension, use, and learning, will ensure a smooth UX that will allow children to focus on the primary purpose of the assessment (Weinerth, 2015).

Therefore, in this doctoral dissertation, I placed significant emphasis on evaluating existing UX approaches to address the target demographic (school aged children between 6 and 12 years).

2. Research approach

2.1. Motivation

As the use of technology in education increases, changes are becoming more apparent, and teachers are more committed to using technology in student assessments. However, clear procedures for evaluating the fairness and validity of e-assessments for all students have yet to be established. Furthermore, integrating technology into assessment practices has not resulted in significant changes in this regard but has added yet another layer of complexity to the future of education. To improve the reliability of e-assessments and reduce the influence of HCI obstacles, it is necessary to observe children's UX while they use applications. By ensuring a high level of UX with respect to interface usability, its ease of comprehension, ease of use, and ease of learning, children are able to focus on the primary purpose of the assessment (Weinerth, 2015). Furthermore, it is imperative for a comprehensive e-assessment system to consider a diverse array of user attributes, such as age, cultural background, and any additional specifications (Reich & Petter, 2009). It is crucial to consider age when examining these characteristics, as a significant portion of current research on UX and its underlying techniques has focused primarily on adult users (Fitton et al., 2014).

Therefore, the core motivation behind this doctoral dissertation was a desire to improve our understanding of how best to evaluate interactive systems designed for young children (Markopoulos & Bekker, 2003). When my coauthors and I began conducting the studies included in my doctoral research, we realized that there was a need for better knowledge about how to evaluate e-assessments, specifically those administered to young children. A systematic investigation, conducted in Chapter 2, revealed a lack of interaction design methodologies that can accurately measure the constructs underlying UX for the age group of 6- to 12-year-olds. Additionally, we recognized the need for techniques that can be administered before or after the assessment to avoid interrupting the children's performance during the assessment itself. We initially chose questionnaires as the primary tool for assessing the children's UX. However, it became apparent that the limited number of scale designs, which are crucial elements of any questionnaire intended for children (but especially for the age group we looked at), needed more rigorous evaluation.

Consequently, we decided to investigate the scale design as an initial measure. Examining numerous biases in these studies prompted us to adopt an expert viewpoint to (1) address the challenges comprehensively and (2) contribute to integrating knowledge in a multidisciplinary field. This allowed us to gather insights from various fields of expertise, such as psychology, education, and child development, thus enhancing the overall quality and validity of our research.

2.2. Background

In this doctoral dissertation, my coauthors and I collaborated with two partners who helped us by providing access to schools, research participants, and e-assessment platforms. LUCET, the first partner, is a research and transfer center specializing in large-scale assessment and testing in technology-rich multilingual learning settings (LUCET, 2023). LUCET's most prominent and resource-intensive commissioned research consists of the establishment, improvement, and assurance of Épreuves Standardisées (ÉpStan) (LUCET, 2023), a Luxembourg school monitoring program that supports evidence-based education policy. LUCET delivers timely and policy-relevant information to national educational stakeholders while compiling a unique and extraordinarily comprehensive longitudinal database chronicling students' competence profiles and their routes through school—and possibly life. LUCET's other mission-oriented research and assessment programs include the national analysis and reporting of international large-scale studies (*About PISA*, n.d.), cognitive and language testing, university admissions testing, and student course evaluations. LUCET's online assessment system, OASYS, administers several of the abovementioned, oftentimes high-stakes examinations. LUCET coordinates, facilitates, and disseminates nationally grounded educational research at the University of Luxembourg and its educational measurement research. We have worked with LUCET on various projects (e.g., an observational study in Luxembourgish schools, a heuristic evaluation workshop) (LUCET, 2023).

The second partner, the Direction de l'évaluation, de la prospective, et de la performance (DEPP) (in English: Direction of evaluation, prospective, and performance) (*Direction de l'evaluation de la prospective et de la performance (DEPP)*, n.d.), has assigned its primary activity to evaluating and measuring skills in education and training. It contributes to assessing the policies conducted by the Ministry of National Education. The complex activity requires that the measurement of the skills be as fair as possible to assess students' actual levels of mastery. Therefore, the collaboration with the Human-Computer Interaction Research Group of the University of Luxembourg has begun to address usability and UX concerns with their tools. In an initial workshop with the main stakeholder of DEPP, the following overall goal of the collaboration was specified: *to enhance the quality of assessment by improving the assessment tools and by creating equal chances for all its users.* The cooperation, exchanges, and valorization of the research outcomes were planned accordingly to communicate the results of the studies. As DEPP is part of the FLIP+ e-assessment (*Flip+ E-Assessment Community*, n.d.), an association that shares its knowledge and experiences, technology development, and content related to e-assessment, the outcomes of the studies were also shared and valorized there. The primary output for DEPP consisted of recommendations on how to improve the test design of their current e-assessment application.

2.3. Overarching research objectives and contributions

In this dissertation, we examined the challenges involved in UX in primary education e-assessments for children aged 6 to 12. Implications for a better interpretation of UX evaluation methods for children were considered a necessary first step for achieving this objective. We formulated the following research objectives:

The first objective was to provide an overview of recent developments and empirical methodologies in the CCI field, as doing so would enable us to ultimately select suitable methods for investigating the UX of children during e-assessment. To this end, we conducted a systematic literature review (Chapter 2).

The second objective was to investigate the cognitive thought processes underlying three commonly used survey and scale designs. The scales were aimed at measuring children's self-reported accounts of their UX. To this end, we conducted a cognitive interview study and incorporated different survey designs into a large-scale questionnaire we administered to children (Chapter 3).

The third objective included experts' perspectives on how to design fair and user-friendly e-assessments. To this end, we developed 10 usability/UX heuristics on the basis of the literature (i.e., previous evidence-based guidelines), validated with an expert survey and a heuristic evaluation workshop. Additionally, we made recommendations on how to improve the QRR methodology, which we used to develop the heuristics (Chapter 4).

The main contribution of all three research papers in this doctoral dissertation is **methodological**.
According to Wobbrock & Kientz (2016), a methodological contribution adds to or refines how researchers or practitioners carry out their work.
Our methodological contributions thereby include:

- A systematic evaluation of recent research methods to give designers and researchers from different domains a unified overview of CCI design methodologies (Chapter 2)
- An investigation of child-appropriate survey methodologies to analyze children's UX with e-assessments (Chapter 3)
- The designing of UX/usability heuristics to evaluate e-assessments administered to children (Chapter 4)

Additionally, we made applied contributions to the existing research by:

- Formulating several recommendations to establish a current framework for developing usability heuristics
- Providing insights from cognitive interviews and digital questionnaires on how to improve pictorial response scales for children

These applied contributions are essential for any practitioners in the field.

In Chapter 5, we discuss the contributions and their output in more detail.

2.4. Research Framework

This doctoral dissertation pertains to the fields of psychology and Human-Computer Interaction (HCI), specifically focusing on the area of CCI. Additionally, I have integrated information from various disciplines, including, for instance, from the educational sciences. From an epistemological perspective, a constructivist stance should primarily be adopted when looking at this work, as such a stance maintains that different people may construct meaning in different ways, even about the same phenomenon (Ültanir, 2012). As people form personal interpretations of their experiences, which are focused on specific items or entities, the researcher seeks to explore the diverse and numerous interpretations rather than simplifying them into a limited number of categories or concepts. The objective of the research is to heavily depend on the participants' perspectives about the subject under investigation. "UX occurs in and is dependent on the context in which the artefact is experienced" (Lallemand & Koenig, 2020) (p.2) and has therefore often a focus on qualitative approaches (Creswell, 2009). However, we emphasize the importance of triangulation methods for fully understanding the intricacies of UX (Pettersson et al., 2018). Therefore, qualitative and quantitative viewpoints and including different stakeholder roles are necessary to reconcile scientific rigor with practical relevance to investigate the area of UX. Whereas our studies did not investigate all possible approaches for evaluating children's UX in the context of e-assessment, we did integrate many research viewpoints, and we emphasize the UX and usability to improve system design so that children will have a good experience while taking the assessment (Weinerth, 2015). The complexity that hails from this stance requires different disciplinary perspectives and methods to be taken into account in order to develop a comprehensive understanding. Although the current work may prioritize research methods above design methods, we did not explicitly distinguish between them (Sanders, 2008). In Research Paper 2, we investigated hedonic components (Hassenzahl, 2001) (e.g., pleasure measured with the Smileyometer Scale (Read & MacFarlane, 2006)) that encourages users to use the e-assessment tool. In research Paper 3, in addition to ISO 9241's usability qualities of effectiveness and efficiency (International Organization for Standardization (ISO), 2018), we addressed the following UX concepts: learnability (i.e., how easy it is for users to accomplish basic tasks the first time they encounter the design), memorability (how easily users can reestablish proficiency after a period of not using the design), and errors (partly covered by the ISO's effectiveness definition, errors can help users avoid mistakes and help users recover, which might affect the outcomes' validity) (Nielsen, 1994, 2012). Additionally, we examined UX accessibility (Morville, 2004) in Research Paper 3 because e-assessment systems should provide all students with fair chances. The decision to examine UX from the perspectives of both children and experts was motivated by the challenges of conducting research with children, for instance, extreme positive response (Read, 2008), particularly in the present context as outlined in detail in Sections 1.3 and 2.1. The perspective of experts can enhance the validity of present procedures and provide insights into the limitations of these approaches, thus contributing to the synthesis of knowledge in a multidisciplinary domain. However, future researchers should work to create a better balance of power relations by experimenting with new methods in the area of research-led and participatory mindset that would give children a stronger voice (Druin, 2002; Sanders, 2008).

The results led to the establishment of a body of scientific knowledge that has practical implications for professionals seeking to understand and advance the UX and usability of e-assessment systems.

2.5. Structure of the dissertation

Chapter 1 provides a high-level overview of where the field of e-assessment currently stands. We summarize the main problems that have emerged from recent developments in e-assessments, focusing on the UX of children. Additionally, we report the challenges of designing e-assessments for children compared with the adult population. At the end of the chapter, we share our research objectives and motivation.

Chapter 2 provides the most recent approaches to children's interaction design by systematically reviewing 272 complete publications. Through our investigation, we comprehensively capture and explain the latest empirical approaches used in CCI. We identify several trends in technologies (e.g., VR/AR) and the preference for combining two or more product evaluation methods. Additionally, we highlight the significance of novel combinations within the extensive spectrum of research- and design-led methods, such as combining various stakeholder participation roles. Finally, we stress the significance of providing additional age-appropriate approaches and activities for younger children. The findings from this paper helped us select our tools for assessing UX during e-assessments from two sides: self-reported accounts of experience from the children and the experts' perspectives on the system.

In Chapter 3, we look closely at three widely used design scales (i.e., the Smileyometer Scale [SOS], the Thumbs-Up Scale [TUS], and the Nodding Heads Scale [NHS]) to evaluate children's UX with the e-assessment platforms. This paper focuses, in particular, on the cognitive challenges of designing scales for young children, as this field remains under-researched and forms the necessary first step in evaluating the suitability of a scale. We derived design recommendations and ideas for future scales for assessing children’s UX.

In Chapter 4, we present heuristics designed for investigating the UX of children with e-assessments. The heuristics are based on experts' knowledge and add to the children's self-reported experience by (a) overcoming biases and limitations as reported in Chapter 3 and (b) providing a holistic view of the UX for designing future e-assessment systems. The heuristics may be used to safeguard the UX of children while taking e-assessments so they can focus on the main work at hand.

In Chapter 5, we synthesize the results from the three research papers in a General Discussion. Moreover, we outline the exact research contributions and explain how the studies made these contributions. We summarize the limitations and offer an outlook on future research on e-assessments for children.

2.6. Associated content and publications

Chapter	Associated publications	Publication status
2	Lehnert, F., Niess, J., Lallemand, C., Markopoulos, P., Fischbach, A., & Koenig, V. (2022). Child–Computer Interaction: From a systematic review towards an integrated understanding of interaction design methods for children. *International Journal of Child-Computer Interaction*, *32*, 100398. https://doi.org/10.1016/j.ijcci.2021.100398	Published
3	Lehnert, F., Koenig, V., Doublet, S., Lallemand, C., Fischbach, A. & Sacré, M. (in preparation for resubmission). Make Me Smile: Exploring Cognitive Thought Processes in Pictorial Scale Designs for Children.	Submitted
4	Lehnert, F., Doublet, S., Koenig, V., Lallemand, C. & Gavin, S. (submitted). Designing Usability/UX Heuristics to Evaluate Educational E-assessment with Children.	Submitted

Additionally, the following related publications and presentations were not included herein:

Lehnert, F.K. User Experience Challenges for Designing and Evaluating Computer-Based Assessments for children. In *Proceedings of IDC '19, the 18th ACM International Conference on Interaction Design and Children* (2019, June 12) Boise, US.

Lehnert, F. K., Lallemand, C., Fischbach, A., & Koenig, V. User Experience Challenges of Children for Designing Computer-Based Assessments. Poster presented at *LuXera conference*. *November 2019. Belval Campus, University of Luxembourg.*

Lehnert, F. K., Lallemand, C., Fischbach, A., & Koenig, V. Experimenter effects In Children Using the Smileyometer Scale. Paper presented at the *PhD conference in Social Science.* 19 November 2020, Esch-sur-Alzette, Luxembourg.

Lehnert, F. K., Lallemand, C., Fischbach, A., & Koenig, V. How do pupils experience Technology-Based Assessments? Implications for methodological approaches to measuring the User Experience based on two case studies in France and Luxembourg. Presentation at the *LuXera conference*. November 2020. Belval Campus, University of Luxembourg.

Lehnert, F. K, Lallemand, C., Fischbach, A., & Koenig, V. User experience challenges for technology-based assessments. Presentation at the *4th FLIP+ Annual Event*, June 2021.

3. References Chapter 1

About PISA. (n.d.). [OECD]. Retrieved February 6, 2024, from https://www.oecd.org/pisa/aboutpisa/.

Ashton, H., & Thomas, R. (2006). Bridging the gap between assessment, teaching and learning. In *10th CAA International Computer Assisted Assessment Conference: Proceedings of the Conference on 4th and 5th July* (pp. 27-38).

Baddeley, A., & Hitch, G. J. (1974). Working Memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47–89). Academic Press.

Bargas-Avila, J. A., & Hornbæk, K. (2011). *Old wine in new bottles or novel challenges: A critical analysis of empirical studies of user experience*. 2689–2698.

Bennett, R. E. (1998). *Reinventing Assessment. Speculations on the Future of Large-Scale Educational Testing. A Policy Information Perspective.*

Bødker, S. (2006). *When second wave HCI meets third wave challenges*. 1–8.

Bødker, S. (2015). Third-wave HCI, 10 years later—Participation and sharing. *Interactions*, *22*(5), 24–31.

Bunderson, C. V., Inouye, D. K., & Olsen, J. B. (1988). The four generations of computerized educational measurement. *ETS Research Report Series*, *1988*(1), i–148.

Burns, T., & Gottschalk, F. (2019). *What do we know about children and technology?* OECD. file:///Users/florencekristin.lehn/Desktop/PhD/Literature/Books/OECD-what%20we%20know%20about%20children%20and%20technology.pdf

Clariana, R., & Wallace, P. (2002). Paper–based versus computer–based assessment: Key factors associated with the test mode effect. *British Journal of Educational Technology*, *33*(5), 593–602.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed). Sage Publications.

Daly, C., Pachler, N., Mor, Y., & Mellar, H. (2010). Exploring formative e-assessment: Using case stories and design patterns. *Assessment & Evaluation in Higher Education*, *35*(5), 619–636.

Dermo, J. (2007). *Benefits and obstacles: Factors affecting the uptake of CAA in undergraduate courses*.

Direction de l'evaluation de la prospective et de la performance (DEPP). (n.d.). Retrieved February 6, 2024, from https://www.education.gouv.fr/direction-de-l-evaluation-de-la-prospective-et-de-la-performance-depp-12389

Donovan, J., Mader, C., & Shinsky, J. (2007). Online vs. Traditional course evaluation formats: Student perceptions. *Journal of Interactive Online Learning*, *6*(3), 158–180.

Druin, A. (Ed.). (1999). *The design of children's technology*. Morgan Kaufmann Publishers.

Druin, A. (2002). The Role of Children in the Design of New Technology. *Behaviour and Information Technology*, *21*(1), 1–25.

Farrell, T., & Rushby, N. (2016). Assessment and learning technologies: An overview. *British Journal of Educational Technology*, *47*(1), 106–120.

Fitton, D., Bell, B., Read, J. C., Iversen, O., Little, L., & Horton, M. (2014). Understanding teen UX: building a bridge to the future. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems* (pp. 79–82).

Flip+ e-assessment Community. (n.d.). Retrieved February 6, 2024, from https://flip-plus.org

Gilbert, L., Whitelock, D., & Gale, V. (2011). *Synthesis report on assessment and feedback with technology enhancement*.

Harrison, S., Tatar, D., & Sengers, P. (2007). *The three paradigms of HCI*. 1–18.

Hassenzahl, M. (2001). The effect of perceived hedonic quality on product appealingness. *International Journal of Human-Computer Interaction*, *13*(4), 481–499.

Hassenzahl, M. (2018). The thing and I: understanding the relationship between user and product. *Funology 2: From Usability to Enjoyment*, 301–313.

Hassenzahl, M., & Tractinsky, N. (2006). User experience-a research agenda. *Behaviour & Information Technology*, *25*(2), 91–97.

Hollender, N., Hofmann, C., Deneke, M., & Schmitz, B. (2010). Integrating cognitive load theory and concepts of human–computer interaction. *Computers in Human Behavior*, *26*(6), 1278–1288.

Hornby, W. (2003). Assessing using grade-related criteria: A single currency for universities? *Assessment & Evaluation in Higher Education*, *28*(4), 435–454.

Hourcade, J. P. (2015). *Child-Computer Interaction*.

Huta, V., & Ryan, R. M. (2010). Pursuing pleasure or virtue: The differential and overlapping well-being benefits of hedonic and eudaimonic motives. *Journal of Happiness Studies*, *11*, 735–762.

Huta, V., & Waterman, A. S. (2014). Eudaimonia and its distinction from hedonia: Developing a classification and terminology for understanding conceptual and operational definitions. *Journal of Happiness Studies*, *15*, 1425–1456.

International Organization for Standardization (ISO). (2018). *ISO 9241-11:2018: Ergonomics of human-system interaction—Part 11: Usability: Definitions and concepts.* International Organization for Standardization. https://www.iso.org/standard/63500.html

International Test Commission. (2006). International guidelines on computer-based and internet-delivered testing. *International Journal of Testing*, *6*(2), 143–171.

JISC, E., & Team, L. (2010). Effective assessment in a digital age-a guide to technology-enhanced assessment and feedback. *Technology Enhanced Assessment*, 26–28.

Joint Information Systems Committee. (2007). Effective practice with e-assessment. *Higher Education Funding Council for England. Retrieved from Http://Www. Jisc. Ac. Uk/Media/Documents/Themes/Elearning/Effpraceassess. Pdf*.

Jordan, S. (2013). E-assessment: Past, present and future. *New Directions*, *9*(1), 87–106.

Kozma, R. (2009). Transforming education: Assessing and teaching 21st century skills. *The Transition to Computer-Based Assessment*, *13*.

Kyllonen, P. C. (2009). New constructs, methods, and directions for computer-based assessment. *The Transition to Computer-Based Assessment*, 151–156.

Lallemand, C. (2015). *Towards consolidated methods for the design and evaluation of user experience*.

Lallemand, C., Gronier, G., & Koenig, V. (2015). User experience: A concept without consensus? Exploring practitioners' perspectives through an international survey. *Computers in Human Behavior*, *43*, 35–48.

Lallemand, C., & Koenig, V. (2017). *How Could an Intranet be Like a Friend to Me? Why Standardized UX Scales Don't Always Fit*. 9–16.

Lallemand, C., & Koenig, V. (2020). Measuring the contextual dimension of user experience: development of the user experience context scale (UXCS). In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society* (pp. 1-13).

Lehnert, F. K., Niess, J., Lallemand, C., Markopoulos, P., Fischbach, A., & Koenig, V. (2022). Child–Computer Interaction: From a systematic review towards an integrated understanding of interaction design methods for children. *International Journal of Child-Computer Interaction*, *32*, 100398. https://doi.org/10.1016/j.ijcci.2021.100398

Llamas-Nistal, M., Fernández-Iglesias, M. J., González-Tato, J., & Mikic-Fonte, F. A. (2013). Blended e-assessment: Migrating classical exams to the digital world. *Computers & Education*, *62*, 72–87.

LUCET. (2023). *Épreuves Standardisées*. http://www.epstan.lu

Mackenzie, D. (2003). *Assessment for E-Learning: What are the features of an ideal e-assessment system?*

Markopoulos, P., & Bekker, M. (2003). On the assessment of usability testing methods for children. *Interacting with Computers*, *15*(2), 227–243. https://doi.org/10.1016/S0953-5438(03)00009-2

Marriott, P. (2009). Students' evaluation of the use of online summative assessment on an undergraduate financial accounting module. *British Journal of Educational Technology*, *40*(2), 237–254.

Maslow, A. H. (1958). *A Dynamic Theory of Human Motivation.*

McCarthy, J., & Wright, P. (2004). Technology as experience. *Interactions*, *11*(5), 42–43.

Mekler, E. D., & Hornbæk, K. (2016). *Momentary pleasure or lasting meaning? Distinguishing eudaimonic and hedonic user experiences*. 4509–4520.

Mekler, E. D., & Hornbæk, K. (2019). *A framework for the experience of meaning in human-computer interaction*. 1–15.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81.

Morville, P. (2004). *User Experience Design*. http://semanticstudios.com/user_experience_design/

Nielsen, J. (1994). *Usability engineering*. Morgan Kaufmann.

Nielsen, J. (2012). Usability 101: Introduction to Usability. *Usability 101: Introduction to Usability*. https://www.nngroup.com/articles/usability-101-introduction-to-usability/

Osuji, U. S. (2012). The use of e-assessments in the Nigerian higher education system. *Turkish Online Journal of Distance Education*, *13*(4), 140–152.

Parshall, C. G., Davey, T., & Pashley, P. J. (2000). Innovative item types for computerized testing. *Computerized Adaptive Testing: Theory and Practice*, 129–148.

Pekrun, R. (2011). Emotions as drivers of learning and cognitive development. In *New perspectives on affect and learning technologies* (pp. 23–39). Springer.

Pellegrino, J., Chudowsky, N., & Glaser, R. (2001). The nature of assessment and reasoning from evidence. *Knowing What Students Know: The Science and Design of Educational Assessment*, 37–54.

Pettersson, I., Lachner, F., Frison, A.-K., Riener, A., & Butz, A. (2018). A Bermuda Triangle?: A Review of Method Application and Triangulation in User Experience Evaluation. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–16. https://doi.org/10.1145/3173574.3174035

Piaget, J. (1955). *The Construction of Reality in the Child*. NYBasic Books.

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

Redecker, C., & Johannessen, Ø. (2013). Changing assessment—Towards a new assessment paradigm using ICT. *European Journal of Education*, *48*(1), 79–96.

Reich, K., & Petter, C. (2009). eInclusion, eAccessibility and design for all issues in the context of European computer-based assessment. *The Transition to Computer-Based Assessment. New Approaches to Skills Assessment and Implications for Large-Scale Testing*, 68–73.

Ridgway, J., McCusker, S., & Pead, D. (2004). *Literature review of e-assessment.* Futurelab.

Rolim, C., & Isaias, P. (2019). Examining the use of e-assessment in higher education: Teachers and students' viewpoints. *British Journal of Educational Technology*, *50*(4), 1785–1800.

Sanders, L. (2008). ON MODELINGAn evolving map of design practice and design research. *Interactions*, *15*(6), 13–17. https://doi.org/10.1145/1409040.1409043

Sim, G., Read, J. C., Holifield, P., & Brown, M. (2007). *Heuristic Evaluations of Computer Assisted Assessment Environments*. 896–905.

Skinner, B. F. (1958). Teaching Machines: From the experimental study of learning come devices which arrange optimal conditions for self-instruction. *Science*, *128*(3330), 969–977.

St-Onge, C., Ouellet, K., Lakhal, S., Dubé, T., & Marceau, M. (2022). COVID-19 as the tipping point for integrating e-assessment in higher education practices. *British Journal of Educational Technology*, *53*(2), 349–366.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285.

Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*, 261–292.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251–296.

Taras, M. (2002). Using assessment for learning and learning from assessment. *Assessment & Evaluation in Higher Education*, *27*(6), 501–510.

Thüring, M., & Mahlke, S. (2007). Usability, aesthetics and emotions in human–technology interaction. *International Journal of Psychology*, *42*(4), 253–264.

Tubaishat, A., Bhatti, A., & El-Qawasmeh, E. (2006). ICT experiences in two different Middle Eastern universities. *Issues in Informing Science & Information Technology*, *3*.

Ültanir, E. (2012). An epistemological glance at the constructivist approach: Constructivist learning in Dewey, Piaget, and Montessori. *International Journal of Instruction*, *5*(2).

Weinerth, K. (2015). *How does usability improve computer-based knowledge assessment?* sn.

Whitelock, D., & Brasher, A. (2006). Roadmap for e-assessment: Report for JISC. *Retrieved February*, *2*, 2008.

Whitelock, D., Thorpe, M., & Galley, R. (2015). Student workload: A case study of its significance, evaluation and management at the Open University. *Distance Education*, *36*(2), 161–176.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–177.

Williams, J. B., & Wong, A. (2009). The efficacy of final examinations: A comparative study of closed-book, invigilated exams and open-book, open-web exams. *British Journal of Educational Technology*, *40*(2), 227–236.

Wobbrock, J. O., & Kientz, J. A. (2016). Research contributions in human-computer interaction. *Interactions*, *23*(3), 38–44.

Wongvorachan, T., Bulut, O., Tsai, Y.-S., & Lindner, M. A. (2022). Improving Student Feedback Literacy in e-Assessments: A Framework for the Higher Education Context. *Trends in Higher Education*, *1*(1), 16–29.

Chapter 2:

Child-Computer Interaction: From a Systematic Review towards an Integrated Understanding of Interaction Design Methods for Children

Citation: Lehnert, F., Niess, J., Lallemand, C., Markopoulos, P., Fischbach, A., & Koenig, V. (2022). Child–Computer Interaction: From a systematic review towards an integrated understanding of interaction design methods for children. *International Journal of Child-Computer Interaction*, *32*, 100398. https://doi.org/10.1016/j.ijcci.2021.100398

1. Abstract

Child-Computer Interaction (CCI) is a steadily growing field that focuses on children as a prominent and emergent user group. For more than twenty years, the Interaction Design for Children (IDC) community has developed, extended, and advanced research and design methods for children's involvement in designing and evaluating interactive technologies. However, as the CCI field evolves, the need arises for an integrated understanding of interaction design methods currently applied. To that end, we analyzed 272 full papers across a selection of journals and conference venues from 2005 to 2020. Our review contributes to the literature on this topic by (1) examining a holistic child population, including developmentally diverse children and children from 0 to 18 years old, (2) illustrating the interplay of children's and adults' roles across different methods, and (3) identifying patterns of triangulation in the methods applied while taking recent ethical debates about children's involvement in design into account. While we found that most studies were conducted in natural settings, we observed a preference for evaluating interactive artifacts at a single point in time. Method triangulation was applied in two-thirds of the papers, with a preference for qualitative methods. Researchers used triangulation predominantly with respect to mainstream methods that were not specifically developed for child participants, such as user observation combined with semi-structured interviews or activity logging. However, the CCI field employs a wide variety of creative design methods which engage children more actively in the design process by having them take on roles such as informant and design partner. In turn, we see that more passive children's roles, e.g., user or tester, are more often linked to an expert mindset by the adult. Adults take on a wider spectrum of roles in the design process when addressing specific developmental groups, such as children with autism spectrum disorder. We conclude with a critical discussion about the constraints involved in conducting CCI research and discuss implications that can inform future methodological advances in the field and underlying challenges.

Keywords: Child-Computer Interaction, Interaction Design Methods, Children's and Adults' roles, Triangulation, Systematic Review.

2. Introduction

Children have become regular users of interactive technologies in the last two decades. For instance, a recent survey found that 42% of the children in the United Kingdom between the age of five to seven own a tablet (Burns & Gottschalk, 2019). As children became increasingly relevant as users and study participants for interaction designers and researchers in the 1990s (Druin, 1999), the field of CCI began to grow steadily. A community with multidisciplinary influences emerged (Hourcade, 2015), which emphasizes designing interactive technology for children that supports their development as a key design goal, based on contributions from diverse fields, including developmental psychology, learning science, interaction design, and computer science (Hourcade, 2015). Therefore, researchers in this field stress the importance of better understanding of how research methods can be adapted to address emerging developments (Jensen & Skov, 2005).

Research in CCI focuses on questions concerning the design of interactive technologies for children and the effects that technologies can have on children and society (Hourcade, 2015). An important component of this quest has been the endeavor to develop new methods to design and evaluate how children interact with novel technologies. This quest parallels the evolution of the field, since for CCI to stay relevant, its methods should evolve as the world around children changes (Markopoulos et al., 2008). These methods are often adaptations of methods originally developed for working with the adult population, although several methodological advances have been proposed to specifically address children's needs and capabilities.

CCI researchers have recognized the need to take stock of advances in the field and have produced a number of contributions that review and synthesize earlier research results. For example, Markopoulos et al. (2008) present an overview of evaluation methodologies for children's interactive products in a textbook format; however since its publication, there have been substantial methodological developments, which are reviewed below. Fails et al. (2012) synthesized methodological advice for designers of children's technology, focusing on motivation and practical ideas for including children in the design process. Hourcade's (2015) textbook provides an instructional overview of the CCI field that also includes methods for designing and evaluating interactive technologies for children, although it was not intended as a systematic, comprehensive survey of the related literature. We add to these resources by providing a systematic and up to date literature survey of CCI methodology. Our survey follows a number of literature surveys in the CCI field with different areas of focus. For example, earlier reviews focused on aspects such as the agency of children with autism, children in technology design (Spiel et al., 2019), HCI games for neurodivergent players (Spiel & Gerling, 2021), engaging and educating children in technology-making activities (Norouzi et al., 2019; Ventä-Olkkonen et al., 2019), as well as values and ethical aspects (Kawas et al., 2020; Van Mechelen et al., 2020). Our aim is to provide an overview of interaction design methods that can serve as a resource for designers and researchers in this field. More recently, a review of CCI methods by Tsvyatkova & Storni (2019) shared some of the aims of this paper, but focused more narrowly on methodology papers in order to provide guidance for selecting and adapting methods, techniques, and tools for working with neurotypical children. There is a need to extend the methodological discussion to a broad group of children, reflecting the interest this community has paid over the years on their inclusion (Börjesson et al., 2015).

Recent methodological reviews in the field of human-computer interaction for the adult population have stressed the importance of investigating triangulation patterns in method application to better inform scholars and practitioners of the advantages of combining quantitative and qualitative methods (Pettersson et al., 2018). To the best of our knowledge, triangulation patterns have not been systematically explored by other reviews in the CCI field. This review will therefore pay particular attention to the triangulation of methods.

In addition, we investigate the roles adults take in interaction design methods that include roles for children as well in order to uncover patterns and differences in method application. Children's roles are enduring topics in CCI, and adults' roles are particularly significant for work with children as well. This is the first paper to provide such an integrated overview of the state of the art of interaction design methods for children. We meet this challenge by analyzing 272 full papers reporting empirical studies that have been published over the last 15 years. Our work is guided by the following research questions:

- How did interaction design methods and triangulation patterns with children between the ages of 0 and 18 evolve during the period of 2005 to 2020?
- How are interaction design methods applied across groups of developmentally diverse children and in different contexts?
- How do the different roles of children and adults in the research and design process inform the selection of a specific method or vice versa?
- What are current practices regarding ethical conduct in research?

The overarching goal of this review is to provide guidance for future studies in the field and support methodological advancements. Systematic reviews offer a unique opportunity to analyze the literature on a combined set of questions, rather than conducting repeated surveys for individual questions. Addressing a broad set of questions in a single survey can provide a single coherent overview for the research community. Based on earlier work, we identified gaps in the current literature and synthesized the contributions for this systematic review. Our review contributes to the research in this field by (1) examining a holistic child population, including developmentally diverse children and children from 0 to 18 years old, (2) illustrating the interplay of children's and adults' roles with different methods, and (3) identifying triangulation patterns in method application while taking recent ethical debates regarding children's involvement in design into account. We begin the systematic review with an outline of similar reviews in the field to further motivate our research interest. We then provide the details of our review methodology and its findings. This is followed by a discussion of the implications of our results.

3. Related work

CCI has an established tradition of and appreciation for literature reviews to analyze current trends and guide future developments of the field. To date, literature reviews on a variety of different topics have been published, such as designing for neurodivergent populations (Spiel et al., 2019; Spiel & Gerling, 2021) and engaging children in technology making activities (Norouzi et al., 2019; Ventä-Olkkonen et al., 2019). In this section, we provide an overview of recent literature reviews in CCI, with a particular focus on methodological reviews, that is, literature reviews that analyze the methods applied in the field and related aspects. The reviews described here were selected in a systematic approach. While setting the scope and framing the context of the current review, we searched for similar reviews in the ACM database published in the period from 2005 to 2020. In our search, we used the terms *review*, *survey* or *meta-analysis* together with the terms *child*, *kid*, *teen*, *youth*, *young*, *adolescent*, or *minor*. The search of the ACM full-text collection resulted in 119 records. After screening all records, we identified a variety of review papers addressing current trends in method application in the CCI domain from different angles (Jensen & Skov, 2005; Yarosh et al., 2011; Börjesson et al., 2015; Tsvyatkova & Storni, 2019).

Jensen and Skov's (2005) review focused on research methodologies used in papers reported in IDC proceedings and selected journals between 1996 and 2004. They classified 105 papers on children's technology design into a two-dimensional matrix on research methods (e.g., case studies, field studies, lab experiments) and research purposes (understand, engineer, re-engineer, evaluate and describe). In contrast, we decided to apply a classification system that describes the underlying methodology in more detail (e.g., user observation, semi-structured interviews, activity logging), similar to the review by Börjesson et al. (2015). Börjesson et al. (2015) developed their framework in line with new trends in the participatory design field by incorporating methods such as creative sessions and prototyping, which we did not find in any of the categories defined by Jensen and Skov (2005). Additionally, Jensen and Skov (2005) found that much research has been conducted in natural settings, with a strong focus on field studies. Our review examines how this trend has evolved in the last fifteen years based on an updated classification framework of methods.

A review by Yarosh et al. (2011) examined the values held by the CCI community. They conducted a content analysis of values expressed in full papers published at the ACM IDC conference between 2002 and 2010. They discussed several types of contributions made by IDC papers and examined, inter alia, the role of the child in technology design based on Druin's framework (2002). Druin (2002) defines four roles that children can play in the research process: user, tester, informant, and design partner. Our review adopts a similar understanding of children's roles by analyzing two of the underlying dimensions upon which these role definitions are based: *the relationship to adults* (which we further refer to as feedback from children, which spans from passive to active), and the *goals for inquiry* (which we further refer to as research vs. design-led), see Figure 1, from (Druin, 2002), p.5.

Role of child	Relationship to adults				Relationship to tech			Goals for inquiry		
	indirect	feedback	dialogue	elaborate	ideas	prototype	product	theory	impact	usability
User	x						x	x	x	
Tester	x	x				x		x	x	x
Informant	x	x	x		x	x	x		x	x
Design Partner	x	x	x	x	x	x	x			x

Figure 1. The four roles children may take on in the design of technologies (Druin, 2002).

We extend such research on the role of the child by also examining the role of adults in the design process. In current methodological approaches in the CCI field , such as participatory design, children are given stronger voices in the research and design process. For instance, there has been a shift away from children taking on roles such as participants in user studies, which are typically applied in a user-centered design approach, to more enhanced roles like peer codesigner (Nesset & Large, 2004). However, the importance of adult involvement throughout the research and design process should not be underestimated. The inclusion of caregivers, teachers, or other experts in the field remains critical, especially for younger user groups and developmentally diverse children (Börjesson et al., 2015). Therefore, we stress the importance of investigating the interplay between these different roles taken on by children and adults in the design and evaluation process, as this could potentially inform the selection of appropriate methods.

In addition to identifying many positive aspects in the surveyed papers, Yarosh et al. (2011) also shared a number of suggestions for improving future research. They concluded that the majority of IDC papers target children between the ages of 6 and 12. The need to work with other age groups seems apparent due to the need to design and evaluate technologies in an age-appropriate way. Thus, our review considers a broad age range, from infants aged 0 years to late adolescents aged 18 years. Moreover, Yarosh et al. (2011) further expressed the need for longer-term evaluations, as stated here: “the difficulty of deploying and evaluating a system over a longer period of time cannot be underestimated, however, neither can the importance of doing so” (p. 143) (Yarosh et al., 2011). They further state that short-term evaluations can merely address usability aspects and engagement, rather than long-term effects such as learning. This review will shed light on the *temporal stage of system usage* (anticipated, momentary, episodic, retrospective, cumulative experience) as reported by Roto et al. (2011), *cross-sectional vs. longitudinal evaluation* (at one moment in time or several moments) and the study’s goals (Druin, 2002), taking up Sanders’ (2008) distinction between between a research-led vs. a design-led approach in her evolving maps of design practice and research.

In 2015, Börjesson et al. (2015) published a systematic literature review of methods for developmentally diverse children. They stated, “compared to other groups of developmentally diverse children, children with high-functioning autism between 8 and 12 years old are the ones that are most often actively involved in the design process. The other groups of children are often taking a more passive role, like being observed, both in the requirements, design, and evaluation phase” (p. 136) (Börjesson et al., 2015). Our review focuses on the whole child population from infants to late-stage adolescents and includes developmentally diverse children. With this approach, we can cross-validate methods rather than treat them as distinctive from one child population to the other. Our goal is to inform the development of more accessible and inclusive research methodologies. Therefore, our analysis of methods is based not on children’s biological age but on the type of feedback children provide across the various methods underlying Druin’s (2002) definition of children’s roles. The type of feedback children are asked to provide guides our evaluation of the appropriateness of a method in a given context.

Tsvyatkova and Storni (2019) conducted a review of methods, techniques, and tools in CCI developed/adapted to support children’s involvement in technology development based on 36 methodological papers published between 1996 and 2015. The authors classify methods and techniques in CCI based on the framework by Sanders (2008), which maps them along a continuum from an expert to a participatory mindset. Sanders (2008) uses this continuum to classify design and research approaches, with their underlying methods, on her evolving map. This idea of a participatory continuum harkens back to earlier research in the CCI community, such as the IBF Participatory Continuum Model developed by Read et al. (2002). The letters IBF stand for the three different design contributions a domain expert can make. It begins with the lowest contribution, termed *informant design*, and proceeds to *balanced design* and finally *facilitated design,* the highest level of contribution. The review by Tsvyatkova and Storni (2019) applies the participatory continuum to recent CCI methods, techniques, and tools. However, this review did not include the full framework presented by Sanders (2008), which consists of two continua, from an expert to a participatory mindset and from a research- to design-led selection of methods. As stated before, we believe that children's degree of involvement across different methodologies can provide us with guidance for choosing the right method in a given context. Therefore, we adopted the original framework (Sanders, 2008), introducing the type of children's feedback on the first continuum (similar to the original continuum from an expert to a participatory mindset) and maintaining the second continuum of design vs. research-led approaches. The results obtained by applying this framework have a different base than in Tsvyatkova and Storni's (2019) work because we (1) utilize a different classification scheme for our methods and (2) draw on a selection of empirical studies and case studies rather than a selection of methodological papers. We want to stress the importance of tracing back how the methods we propose are implemented in the field to shape a more realistic picture and state of the art, providing us with the possibility to improve and iterate on our methods.

Keeping the above reviews in mind, one important aspect has not been addressed to our knowledge in the CCI community so far, namely the triangulation of methods. Within the broader human-computer interaction (HCI) research community, a recent review addressed the importance of method triangulation for the adult population (Pettersson et al., 2018). The early roots of triangulation can be found in the different epistemological standpoints taken in social, behavioral, and human science. The interplay of qualitative and quantitative methods can help researchers addressing the same problem from different angles to derive a fuller understanding of it. The application of multiple methods within the same study is known as *concurrent* triangulation and can lead to a more reliable and less biased investigation (Creswell, 2009). A *sequential* triangulation of methods, by first using quantitative methods and later qualitative methods, can lead to a fuller and more holistic picture, highlighting different dimensions of the same phenomenon (Creswell, 2009).

In recent years, we have observed a growing interest in ethics within the CCI community, which are addressed in a recent systematic review by Van Mechelen et al. (2020). In their paper, they explore how and to what extent ethics has been dealt with over the last 18 years. They demonstrate that while ethics is frequently mentioned, the literature remains underdeveloped in a number of areas, including definitions and theoretical foundations, the reporting of formal ethical approval procedures, and the extent to which design and participation ethics has been dealt with (Van Mechelen et al., 2020). As our review seeks to deliver an integrated understanding of interaction design methods based on empirical papers, we were especially interested in how and to what extent formal ethical approval is reported in the CCI literature, in order to provide a realistic picture of the state of affairs.

4. Method

To develop an integrated understanding of interaction design methods focusing on a diverse population of children, triangulation patterns, and the roles of children and adults, we conducted a systematic literature review. This systematic review follows the PRISMA statement, which stands for *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (Moher et al., 2009). We present our detailed methodological approach in this review in 5 main steps.

Step 1: Identification of potentially relevant publications

Search terms. Based on our research questions and inspired by the review by Börjesson et al. (2015), we performed a search query in papers' abstract for the terms *"design method" OR "design process" OR "design approach"* etc. ***AND*** in the title for *child* OR kid* OR adolescent*,* etc. (see Table 1). The characters (“ ”) reflect a search for the exact match of that phrase, whereas (*) specifies a number of unknown characters. For example, when using the expression child*, the search engine results will also contain words such as children or childhood. We limited our second search term to the abstract because we found that searching for the terms in the complete text returned too many false positives, which we would then have to remove later on based on our exclusion criteria. The same was the case for our target group (first search term). After several iterations, we found that searching for the term in the complete text or abstract resulted in too many false positives; therefore, we decided to search for it in the title only.

Table 1. The search query with search terms 1 in title and search terms 2 in abstract

Search Term 1 (Title)	**Boolean operator**	**Search Term 2 (Abstract)**
Child(ren)		Design method(s)
Kid(s)		Design process
Adolescent(s)		Design approach
Minor(s)		Design technique
Youth		Interaction design
Teen		User-centred design
Teenager		User experience
Young	**OR**	Human-centered design
Student		Child-computer interaction
Pupil		Human-computer interaction
		User-system interaction
		Inclusive design
		Universal design

Note: Terms were searched in both their British and American English forms

Database selection. We selected journal papers from the SCOPUS database and ACM Digital Library. A structured database search within SCOPUS identified 6963 potentially relevant papers from January 2005 to July 2020 (when the database search was conducted). We started our systematic review with papers from 2005 onwards, as the earliest literature review we identified in our search for related workcovers studies until 2004 (Jensen & Skov, 2005). After filtering for document type *"article"*, source type *"journal"* and publication stage *"final'',* a total of 2844 references remained. Within the ACM library, we identified a total of 3750 papers. After filtering for proceedings and research articles (to exclude short papers and extended abstracts), 1528 papers were left.

Selection of Journals and Conferences. We included 7 top-level journals (based on the selection by (Jensen & Skov, 2005)) plus the International Journal of Child-Computer Interaction. In line with arguments by Jensen, J. & Skov, M. (2005) , we find that the pool of included journal papers provides a solid base for our review given the number of papers and the journals' high-quality peer review process. We further narrowed the scope to two suitable target conferences (IDC, CHI) based on their relevance for our study aim. These two venues seemed to be the dominant venues for our target audience because they returned the most hits for our search query. We excluded the International Conference on Human-Computer Interaction (HCII) proceedings, for example, because they are only reviewed based on abstracts, and we wanted to focus on peer-reviewed venues. We also excluded education technology conferences such as the International Conference on Education Technology and Computers (ICETC) proceedings, because these communities do not focus specifically on developing methodologies for interaction design. In addition, the choice of scope was made in order to retain a manageable corpus of papers.

The PRISMA flowchart (Figure 2) presents the overall distribution of publications per journal and conference venue. We counted a total of 829 papers after checking for duplicates. Additional false positives (e.g., no full paper) were excluded at later stages.

Journals:

- Transactions on Human-Computer Interaction, (ToCHI) ACM
- International Journal of Human-Computer Studies, (IJHCS) Elsevier
- International Journal of Human-Computer Interaction, (IJHCI) LEA
- Behaviour and Information Technology, (BIT) Taylor & Francis
- Interacting with Computers (IwC), Elsevier

- Personal and Ubiquitous Computing, (PUC), Springer-Verlag
- International Journal of Child Computer Interaction (IJCCI), Elsevier

Conference proceedings:

- Interaction Design and Children Conference (IDC), ACM
- Conference on Human Factors in Computing Systems (CHI), ACM

Step 2: Definition of scope and procedure

We conducted two rounds of exclusion. The first round aimed to quickly exclude papers based on formal criteria, whereas the second round narrowed the scope based on content.

Exclusion criteria 1st round. We double-checked for the presence of false positives in our selection based on the paper format. We therefore excluded papers **(1)** that were not full papers. In the Interaction Design and Children Conference proceedings, the distinction between short and full papers is sometimes blurred. According to the paper submission guidelines of the 2019 Interaction Design and Children Conference, papers of 8 pages or more were classified as full papers. **(2)** We further excluded papers that did not include an empirical study, or did not include a clear description of the evaluation process or the study results. Case studies, even with a single child, were included as long as user data were reported in a clear way. Studies based on the author's self-reflection or introspection, without involving children, were excluded. **(3)** Papers were excluded that did not report research with children directly (e.g. research with adults about products for children was excluded). **(4)** Papers reporting studies with participants more than 18 years old were excluded, except if only part of the sample was above age 18 (e.g. 16-24), in which case we included the paper.

Exclusion criteria 2nd round. We excluded papers **(1)** that did not report a new study and solely referred to former ones, **(2)** that relied solely on perceptions or psychological factors, with no tangible or interactive artifact involved. We included, however, early studies that did not evaluate a specific system/technology/tool but described a system/technology/tool authors intended to design in the future.

Procedure. Studies were selected independently by the first two authors based on a screening of titles and abstracts. Miles et al. (1994) (p. 11) emphasize the importance of conclusions being verified, whether by reference back to field notes, achieving "intersubjective consensus" through discussion with colleagues, or replicating findings in another dataset. Thus, the selected papers were cross-checked during two exclusion rounds, and discrepancies were resolved by including the third author to reach a consensus about study inclusion.

Step 3: Identification of relevant publications

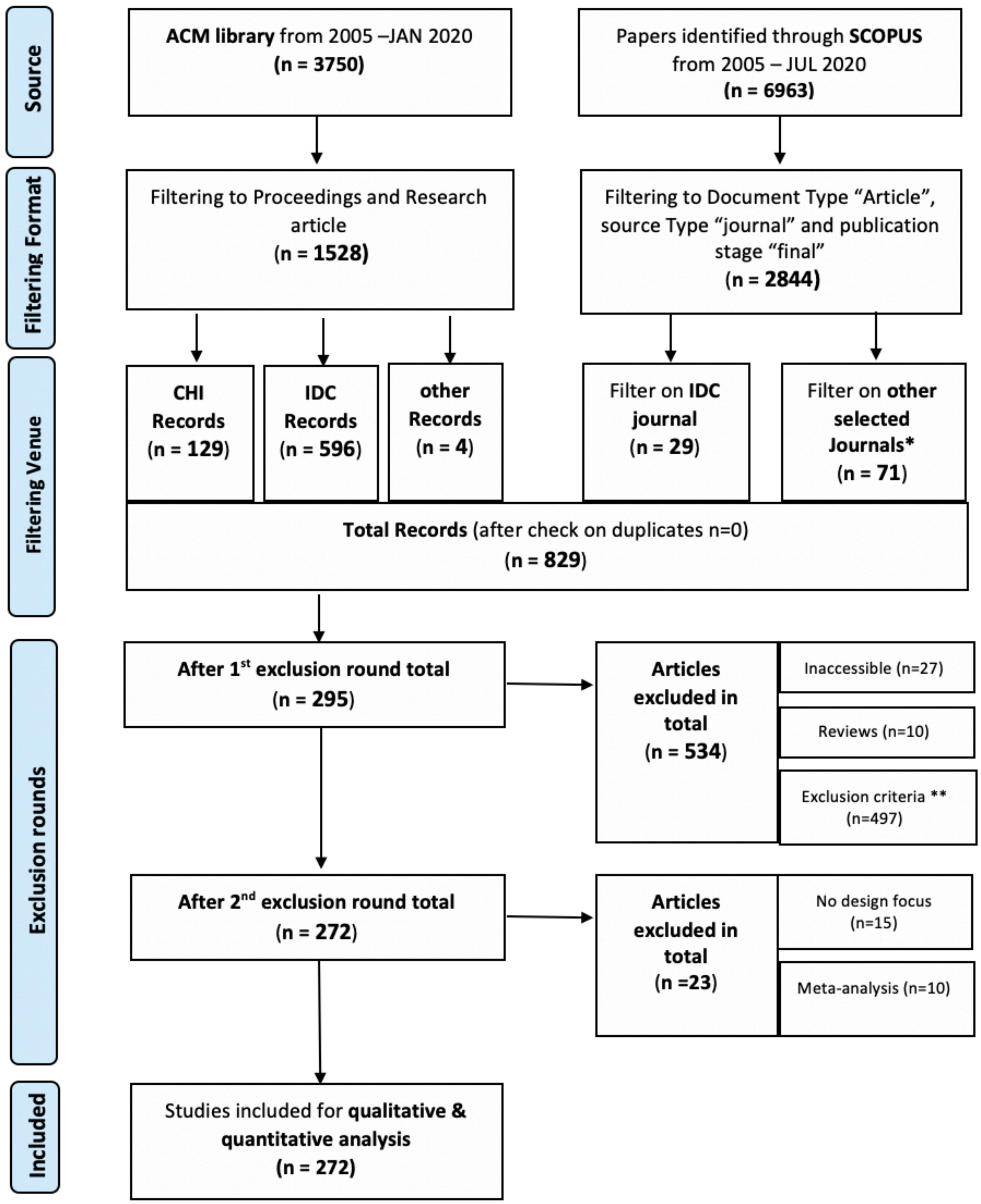


Figure 2. Flow chart of our systematic review adapted from the PRISMA statement (Moher et al., 2009).

Step 4: Categorization of relevant publications

Coding scheme. Our coding scheme was developed in an iterative approach to ensure that the categories were understandable and easy to use. We tested the coding scheme with a group of eight HCI researchers to gain external expert feedback and ensure that coding would not assume specialized skills or prior knowledge. Our final coding scheme used for the analysis consisted of *20 codes*, classified into several **code groups**.

1. **General study setup** codes included the *type of paper* (conference, journal), *type of study* (natural, artificial, environment-independent, combination), *study location*, and *product evaluated*.
2. **Methodology** codes included the *temporal stage of system use* (anticipated, momentary, episodic, retrospective, cumulative experience, and any combination), *longitudinal vs. cross-sectional evaluation*, classification of the *special developmental group, method triangulation, data triangulation (qualitative vs. quantitative), method type* (e.g., user observation, semi-structured interviews, standardized questionnaires, etc.), the *specific methods used* (note that we did not distinguish between methods, techniques and tools, as most of the papers included in this review did not do so) (Börjesson et al., 2015) .
3. **Information about the sample** was also coded, including the *age of the children* (including the upper and lower age bounds) and the *number of participants*. Those codes were applied for all methods separately in cases in which multiple methods were reported in a single paper.
4. **Children's roles** were classified based on Druin's model (2002), focusing on the relationship to adults *(indirect, feedback, dialogue, elaborate*), which we hereafter refer to as *feedback from children* on a continuum from *passive to active*, and the goals for inquiry (*theory, impact, usability*), which we hereafter refer to on a continuum from research to design-led (see Figure 1). We also coded the form of involvement of adults in the research and design process by using Börjesson's (2015) classification to distinguish *adults' roles* (user, proxy, expert, facilitator, support).
5. **Ethical guidelines** focus on *informed consent* and *ethical approval* as mentioned in the papers.

Step 5: Analysis of relevant publications

Procedure. All included papers were coded by the first author. 10% of the papers were double-coded by the first and second authors. The interrater reliability was found to be Kappa = 0.59 ($p < .05$). According to Landis and Koch (1977), this is a moderate level of agreement. However, we found it more important to achieve group consensus by reconciling discrepancies between the coders than to achieve a high inter-rater reliability score (Blandford et al., 2016) (p.93). Therefore, the coders extensively discussed the 10% of double-coded papers on a weekly basis. For the remaining 90%, the first author marked the papers about which she was uncertain and asked the second author for input during their weekly meeting. In cases of disagreement, the third author was brought in, and corresponding adjustments in the coding scheme were made.

Quantitative Analysis. The quantitative analysis of the relevant coded data set was conducted in SPSS v.25.

5. Results

We present the results beginning by charting the temporal trends and perspectives, followed by a description of the child population addressed in terms of age and developmental group. We then elaborate on the context in more detail, including research location, type of study, type of product and longitudinal vs. cross-sectional studies. Subsequently, we examine the type of method and triangulation patterns. We summarize our methodological findings in a graphical overview of interaction design methods for children based on the type of feedback from children and the degree to which the reported studies are design vs. research-led. We conclude with a brief analysis concerning ethics in research with children.

5.1. Temporal perspective and trends

Our review shows an increase of both conference and journal papers between 2007 and 2020 (see Figure 3). Based on our exclusion criteria (see Section 3.4), we did not identify any relevant papers between 2005 and 2006. Furthermore, we only included conference papers archived in the searched databases by July 2020, when our search was executed; therefore, we see a sharp drop in Figure 3. Conference papers seem to be the more popular venue for publications in the field of child-computer interaction. One reason for this development could be the establishment of the Annual Conference on Interaction Design and Children in the Netherlands in 2002. Interestingly, there has been a sharp rise in conference papers in recent years. Before the first exclusion round, 71.8% of papers in our review were from the proceedings of the Interaction Design and Children Conference. Additionally, we included the ACM Conference on Human Factors in Computing Systems (CHI), which is one of the most prestigious conferences in the broader HCI field (15.5% of all papers included in our review are CHI papers).

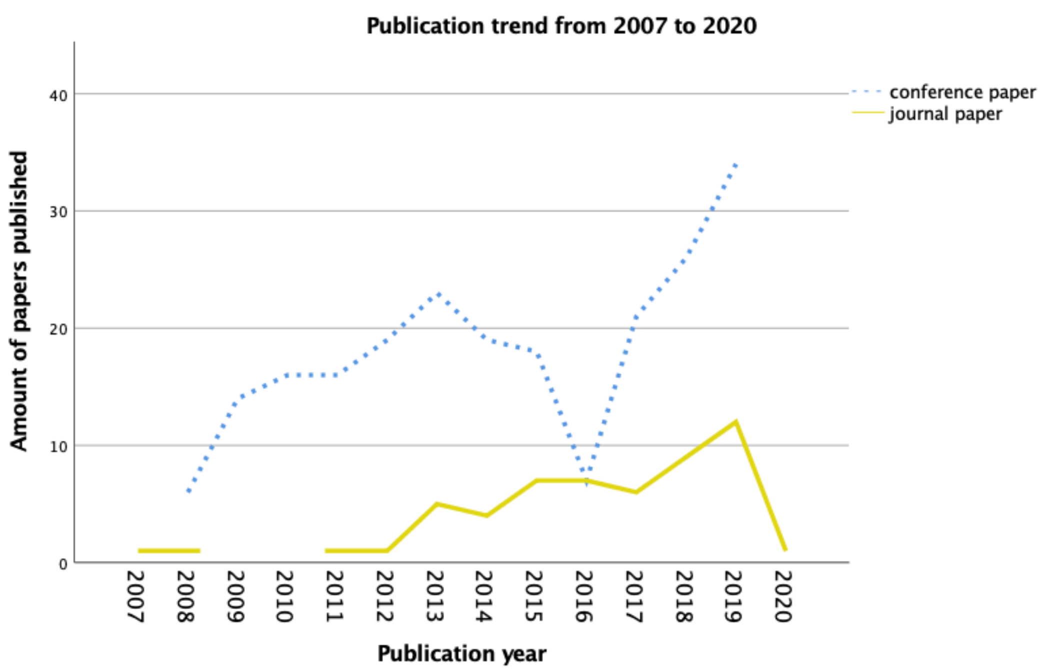
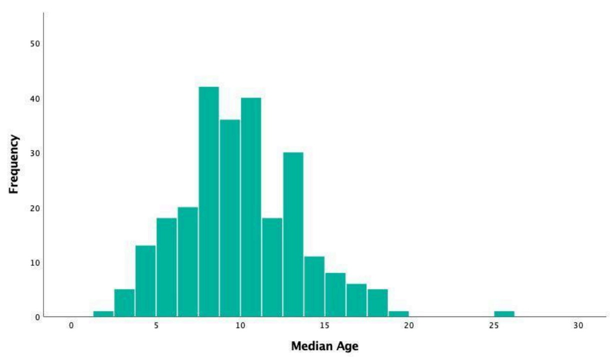


Figure 3. Publication trend for all papers included in the review, from 2007 to 2020 (conference venues until January 2020 and journal papers until July 2020).

5.2. Population of children

The sample size of children participating in the studies ranged from N=1 to N=1031, with an average sample size of 71.03 (SD=115.85). Children from all age groups are represented, with a primary focus on those 7-13 years old (Figure 4 left). We calculated the children's median age (based on the lower and upper age range, because only some papers reported the mean age) M= 9.88 (N=255 papers, N=18,610 child participants overall; SD= 3.53). In 17 papers, the participants' age was not stated.



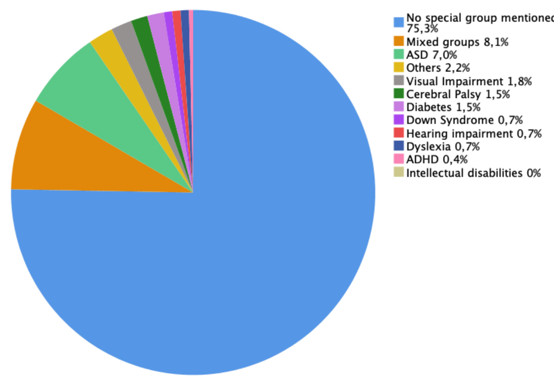


Figure 4 (left). Age of child participants involved in the reviewed studies, across all interaction design and research methods (n=255).

Figure 5 (right). Distribution of child participants across developmental groups, within reviewed papers (n= 272)

Overall, we can state that children participate across all studies and methods. We found that 75.3% of papers did not mention a specific developmentally diverse group (see Figure 5). Autism Spectrum Disorder (ASD) is the most frequently represented group within developmentally diverse children, at 7% of papers. Some papers also report studies with mixed groups of normally developing and developmentally diverse children, accounting for 8.1% of total participants across the reviewed papers.

5.3. Context

Research location. The majority of research studies (37.9%) took place in educational institutions (e.g., schools, preschools). 10.7% took place in a laboratory environment, 7.4% at the participants' homes, 5.1% on the premises of health care providers (e.g., clinic, hospital, rehabilitation center), 4% in museums and 3.7% in other places. We also observed that 12.1% of the papers report studies that were conducted in a combination of locations.

Type of study. We distinguished between a natural setting (field studies, case studies, action research), artificial setting (e.g., laboratory), environment-independent/online setting (e.g., survey research, applied research, basic research, normative writing), or a combination of these. Jensen et al. (2005) applied the same classification in their review, based on earlier work (Wyeth & MacColl, 2010). In our sample of reviewed papers, 55.2% of studies were conducted in a natural setting, 31.3% in an artificial setting, 13.1% in an environment-independent setting, and only 0.4% in a combination of the above. For 13 papers, we could not identify the exact type of study.

Type of product. The most common products studied from 2005 to 2020 were *interactive games* (25.4%), followed by *professional software* (12.1%), *application/mobile phone* (11.8%), *VR/AR* (4.0%), *wearables* (2.6%), *mix of products* (2.6%), *webtools* (2.2%), *websites* (1.1%), *audio/video/TV* (0.7%), *non-digital products* (0.7%) and *other products* (34.7%). We did not further classify this last category due to the heterogeneity in products examined. Examples include a pet robot (Segura et al., 2012), a multimodal collaborative music environment (Zhou et al., 2011), an interactive tabletop (Rick et al., 2011) and a digital clay interface (Follmer & Ishii, 2012). We also classified non-use of a product (n=16), for instance developing an educational toolkit (Wyeth & MacColl, 2010), under *other products*. In Figure 6, it can be seen that products such as VR/AR have arisen more recently in the published papers, whereas interactive games, application/mobile phone and professional software have a longer history.

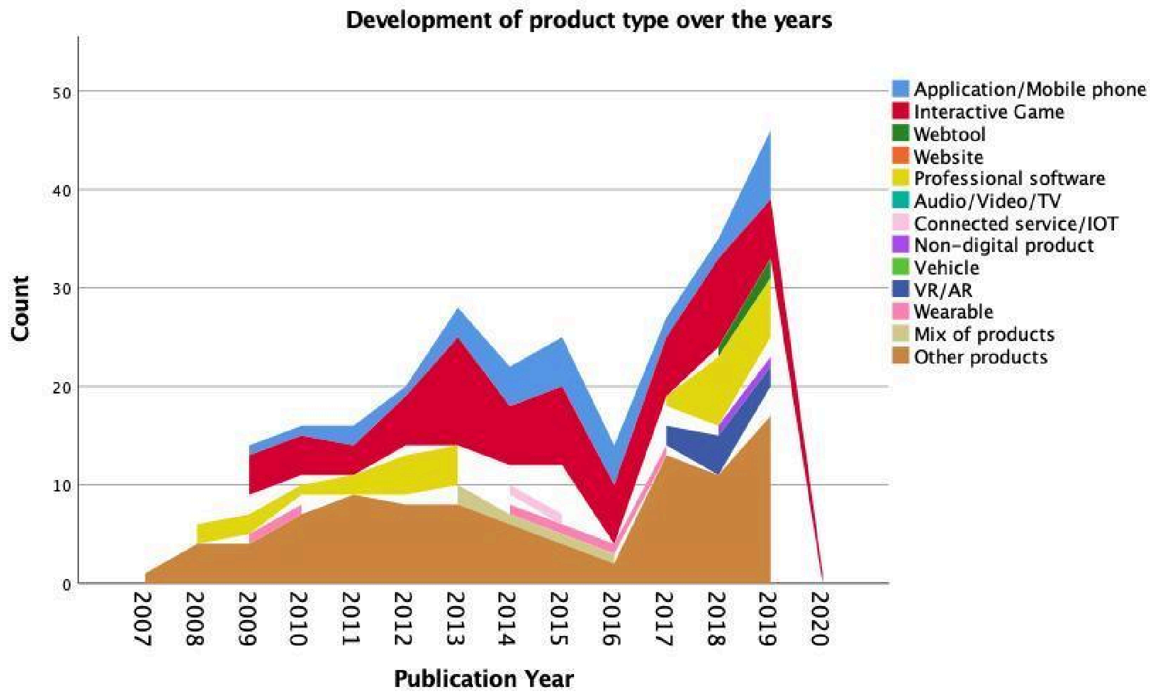


Figure 6. Distribution of product types over time until July 2020.

Examining the distribution of products per developmental group (Figure 7), *interactive games* are applied for all developmental groups except ADHD. The results presented in Figure 7 should be interpreted with caution, as the numbers for several developmental groups are quite low. Moreover, *professional software* is also studied in a number of groups, such as ASD, diabetes patients, as well as the typically developing group (no developmentally diverse group mentioned) and mixed groups. An example of *professional software* is computer-generated audio and visual feedback to encourage vocalizations among children with ASD (Hailpern et al., 2009). Apart from typically developing children, ASD seems to be the most frequent target group in the studied products. In Figure 7, we can also see that a lot of *other products* are used. Moreover, most products were studied in a natural setting: connected service/IOT (100%), wearable (85.7%), VR/AR (70%), interactive games (60.2%). We only found webtools in an artificial setting (80%). Mobile phones/applications and professional software were almost evenly studied in artificial as well as natural settings.

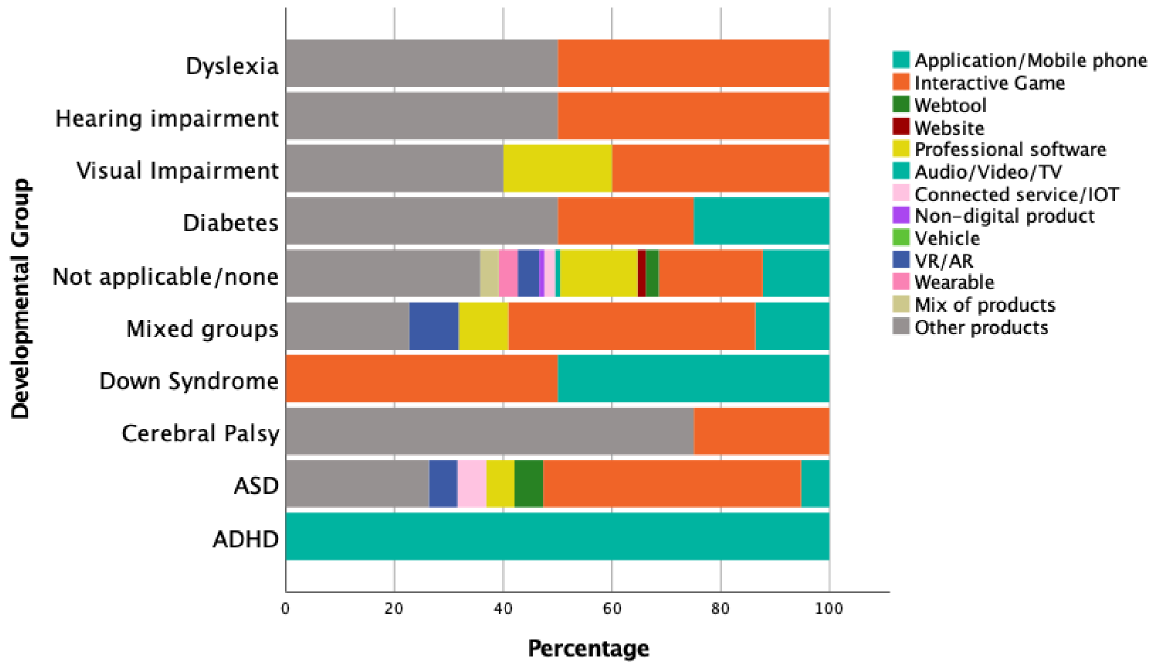


Figure 7. Type of product by group of children.

Longitudinal vs. cross-sectional studies. 58.1% of the studies evaluated a product or system at only one moment in time (cross-sectional studies). Of these, 55.9% were conducted in a natural setting. In 41.9% of cases, a product was evaluated several times (longitudinal studies), 44% of such studies were in a natural setting. In terms of development over time, we see a sharp rise in longitudinal studies in recent years, from only two studies reported in 2008 to nine studies reported in 2016 and 20 studies in 2020.

5.4. Method application

We clustered the research and design methods into predefined categories inspired by the reviews by Börjesson et al. (2015) and Petterson et al. (2018). Börjesson et al. (2015) classified methods and techniques together, arguing that "while there is a fine distinction between methods and techniques we will use these words interchangeably, as most of the papers included in this literature review also do" (p. 80). We decided to merge methods and techniques used in the participatory design field, such as drawing and creating stories with users, under "creative sessions/workshops", similar to Börjesson et al. (2015). In contrast to Börjesson et al. (2015), we decided not to distinguish between pre- and post-interviews but between the exact interview methods used (semi-structured vs. free/open interview), as in Pettersson et al. (2018). We also further specified questionnaires into self-developed questionnaires and standardized questionnaires, as in Pettersson et al. (2018). In contrast to Börjesson et al. (2015) and Pettersson et al. (2018), we added cooperative inquiry as a stand-alone method due to the high number of papers mentioning it as a method and the specific methods/techniques underlying it, such as Sticky noting (Kumar et al., 2018; Woodward et al., 2018). We further decided to classify prototyping as a stand-alone method because we coded it as design-led to a lesser extent compared to the creative sessions (see description of the coding scheme in Section 3.4 point 4). The same was true for brainstorming, which had a stronger research-led focus compared to the creative sessions and was therefore also given a separate method category.

Our coding revealed that the studies reported in the reviewed papers applied various methods, the most popular one being user observation (23.9%). We included in this category field observation and video observation conducted in a laboratory setting. We could identify some observation schemes developed for specific user scenarios, such as outdoor play (Ma et al., 2018). The second most popular method was semi-structured interviewing (16%), followed by self-developed questionnaires (11.7%) and creative sessions and workshops (11.2%), which encompass a variety of sub-methods and techniques (see Table 2 for all percentages and definitions of the methods).

Table 2. Overview of method types and their definitions along with the underlying exact method, technique or tool mentioned by the authors. In the percentage column, we report the frequency of each type of method as a share of all studies. Note: not all reviewed papers mentioned an underlying method, technique or tool and were therefore left blank.

Type of method (with an example reference from our review)	**Definition**	**%**	**Name of exact method, technique or tool** (with an example reference from our review)
User Observation (Hiniker et al., 2017)	"Observation involves watching and noting what happens, and usually takes place in the situation where the technology of interest is or will be used. The focus may be on work or leisure activities, and how the technology supports, hinders or otherwise shapes them, or on people's interactions with the technology." (Blandford et al., 2016)(p. 36)	23.9	Outdoor play observation scheme (OPOS) (Ma et al., 2018), Observational System of Motor Skills (OSMOS) (Landry et al., 2013) , System for Observation of Children's Social Interactions (Raffle et al., 2010)
Semi-structured interview (Scharf et al., 2012)	"In semi-structured interviews many questions (or at least themes) will be planned ahead of time, but lines of enquiry will be pursued within the interview to follow up on interesting and unexpected avenues that emerge." (Blandford et al., 2016)(p. 40).	16	Question-asking protocol (Hudson et al., 2018), this-or-that (Zaman & Abeele, 2010), Rapid desirability testing (Fleck et al., 2018)
Self-developed questionnaire (Fletcher-Watson et al., 2016)	A questionnaire developed by the researchers whose psychometric properties have not yet been tested (unlike for standardized questionnaires, see (Boynton & Greenhalgh, 2004)).	11.7	

Creative sessions/workshops (McRoberts et al., 2019)	Methods and techniques used in participatory design, such as drawing and creating stories with users (Börjesson et al., 2015) and all other creative design activities/workshops not explicitly named as part of the cooperative inquiry method.	11.2	Co-design (Hamidi et al., 2014), storyboarding (Schaper et al., 2018), design critique (Frauenberger et al., 2013), track method (J. C. Read et al., 2014), drawing-telling (Desjardins & Wakkary, 2011), design exposès (Frauenberger et al., 2016), obstructed theatre (Mazzone et al., 2010), persona (Jones et al., 2012), scenario-based design (Vacca, 2019), cool wall (Bowen et al., 2013), blue sky (Bowen et al., 2013), x-factor (Bowen et al., 2013), dragons den (Bowen et al., 2013), group sorter (Soute et al., 2013), context mapping (Van Mechelen et al., 2015), drama-based method (Frauenberger et al., 2017)
Activity logging (Nacher et al., 2018)	Monitoring, recording or tracking user actions such as the target of a click, hover event, current mouse position, etc. and nearly any other on-screen activity (Atterer et al., 2006).	9.9	

Standardized questionnaires (Metatla et al., 2019)	Usage of a previously validated and published questionnaire (Boynton & Greenhalgh, 2004).	5.4	Smileyometer (Leite et al., 2017), Fun Sorter (Sim & Horton, 2012), Again-again Table (Sim & Horton, 2012), Self-assessment Manikin (Metatla et al., 2019), Attrakdiff (Fleck et al., 2018), Intrinsic Motivation Inventories (IMI) (Harms et al., 2013), Kids Game Experience Questionnaire (KidsGEQ) (Martin-Niedecken, 2018), PA Measure-Revised (MPAM_R) (Ma et al., 2018), Playful Experiences Questionnaire (PLEXQ) (Ma et al., 2018)
User testing (Bateman et al., 2018)	Users are invited to complete typical tasks with a product or simply asked to explore it freely while their behaviors are observed and recorded in order to identify design flaws that cause user errors or difficulties. The term is often used interchangeably with usability tests (Bastien, 2010).	4.7	Think aloud (Leduc-Mills & Eisenberg, 2011)
Diary keeping (Sonne et al., 2016)	"Diary studies enable participants to record data in their own time, at particular times of day or when a particular trigger occurs. Diary entries may be more or less structured." (Blandford et al., 2016)(pag. 46)	2.3	Memoline (Sim et al., 2016)

Focus groups (Gray et al., 2015)	"The researcher typically takes a role as facilitator but the main Interactions are between participants, whose responses build on and react to each others'." (Blandford et al., 2016)(p. 45).	2.3	
Cooperative Inquiry (Garzotto, 2008)	In the cooperative inquiry method, adults and children use a broad range of techniques to work together throughout the entire design process to create new technology (Fails et al., 2012).	2.1	Sticky noting (Kumar et al., 2018), kid reporter technique (Schaper et al., 2018), bags of stuff (Yip et al., 2019), Big Props (Yip et al., 2019), line judging (Yip et al., 2019), big paper technique (Woodward et al., 2018), layered elaboration (Yip et al., 2016), mixing ideas (Fails et al., 2019)
Free interview (Liszio & Masuch, 2017)	"A completely unstructured interview is more like a conversation, albeit one with a particular focus and purpose." (Blandford et al., 2016)(p. 40).	2.1	
Prototyping (Alhumaidan et al., 2018)	Prototypes are the "filters that traverse a design space" and the "manifestations of design ideas that concretize and externalize conceptual ideas" (Lim et al., 2008).	2	
Brainstorming (Chu et al., 2015)	Brainstorming is a technique for fostering group creativity in which ideas and thoughts are shared among members spontaneously in order to reach solutions to practical problems (Al-Samarraie & Hurmuzan, 2018).	1	

Physiological measurements (Antle et al., 2019)	Users’ physiological signals can be used to assess their emotional experience while interacting with the product. Physiological measurements are signals which can be measured for living organisms (e.g. heart rate or skin conductance) (Alexandros & Michalis, 2013).	0.8	Eye-tracking (Maqsood et al., 2018), emotion tracker (Cerezo et al., 2019)
Probes (Brule et al., 2016)	(Cultural) probes gather qualitative data based on participatory user self-documentation. Cultural probes are a collection of tools, typically consisting of single-use cameras, user diaries, maps, postcards, or the like—each item added with some instructions on how to use it (Thoring et al., 2013)	0.7	
Card sorting (MoraGuiard & Pares, 2014)	A methodology that can be used to capture users’ mental models of how information is organized in a software interface. For software and website content, the results also lead to suggestions for navigation, menus, and possible taxonomies (Chaparro & Hinkle, 2008)	0.5	Picture Sorting Task (Kory-Westlund & Breazeal, 2019)
Other		3.1	Robotic Intervention Method (Fransen & Markopoulos, 2010), concept mapping (Michaelis & Mutlu, 2019), thought-listing technique (Jones et al., 2012)

Not clear	0.3

5.5. Triangulation patterns

The following analysis of triangulation is based on the definitions of methods outlined in Table 2. We observed that two-thirds of the papers use at least two methods (see Table 3). When calculating the triangulation patterns, we took only the main methods into account (those listed in the left column of Table 2).

Table 3. Number of methods used in each paper (cumulative percentages)

Number of methods reported	1	2	3	4	5	6 or more
Distribution and cumulative percentage	N = 272 (100%)	n = 182 (66.9%)	n = 93 (34.1%)	n = 43 (15.8%)	n =17 (6.2%)	n = 7 (2.5%)

Looking into concurrent triangulation, that means, the combination of different methods mentioned in the left column of Table 2 within the same study, we found the strongest link between user observation and semi-structured interviews (17.0%), followed by user observation and activity logging (12.6%). Less frequent were combinations of user observation and self-developed questionnaires (7.1%), or semi-structured interviews and self-developed questionnaires (5.4%). In 4.3% of cases, we find combinations of three methods in the same study (typically user observation, semi-structured interviews and self-developed questionnaires). In terms of sequential triangulation, we observed a mixed-method approach, that is, a mix of qualitative and quantitative methods, in 37.9% of papers. Qualitative methods alone accounted for more than half of the selected papers (50.7%), while quantitative methods alone were applied much less frequently (9.2%).

Looking at sequential triangulation in different user groups, we see that qualitative methods are more often applied than quantitative methods for most user groups (see Figure 8). However, a mixed-method approach is a common choice with children with hearing impairments, Down syndrome and ADHD, for instance. In the visual impairment and Autism Spectrum Disorder (ASD) groups, we observed a preference for qualitative methods. In research with ASD children, user observation (13 out of 19) and creative sessions (6 out of 19) are the most preferred qualitative methods. Participatory design is also a common approach and is used as an umbrella term for many creative techniques in which children are actively involved in the design process (Nesset & Large, 2004). Although participation seems challenging for some user groups (e.g., children with ASD), new methods are continuously being developed to involve them in the research and design process, for instance the recently developed design critique tool (Frauenberger et al., 2013). In contrast, we observed that methods involving children with cerebral palsy more often involve gathering indirect feedback via methods such as activity logging or user tests (Saturno et al., 2015). In general, we see a preponderance of mixed methods across all groups of children, with visually impaired children as the only exception.

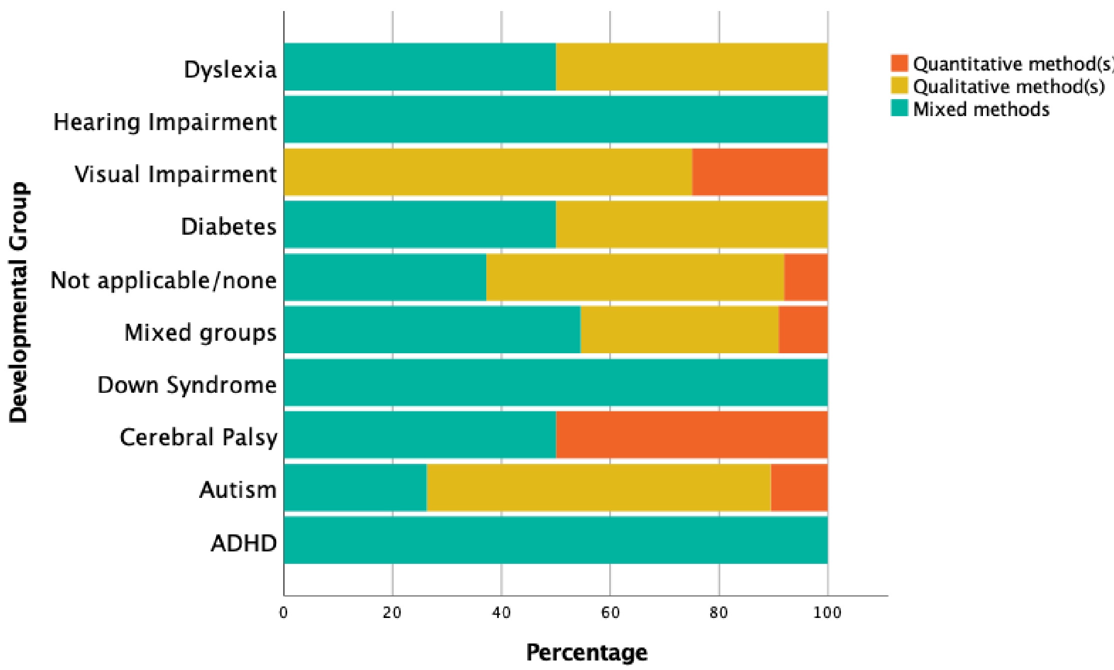


Figure 8. Data triangulation across developmental groups

5.6. An integrated understanding of research and design methods

In Figure 9, we present an overview of all interaction design methods for children based on our selection of papers. We positioned the methods (grey circles) along two axes reflecting the different types of feedback children can provide and research vs. design-led methods (based on the coding scheme described in Section 3.4 point 4). The two dimensions are based on Druin's model (2002), although we focused on children's relationship to adults, which we hereafter refer to as *feedback from children (coded as: indirect, feedback, dialogue, elaborate)*, ranging from *passive to active*, and goals for inquiry (coded as: *theory, impact, usability*), which we hereafter refer to on a continuum from research to design-led (see Figure 1).

The small black circles indicate examples of specific methods and/or techniques named by the authors of the reviewed papers. The size of the circle reflects the relative frequency of the method in all papers (e.g., user observation, at 23.9%, has a bigger circle compared to semi-structured interviews, at 16%). User testing ranges from indirect feedback to verbal feedback, as different papers described different types of feedback provided by children. For methods such as creative sessions, children were mostly reported to be actively involved in the design process and therefore coded as having engaged in either dialogue or elaboration within the feedback dimension. For probes-based methods (where participants typically engage with probe materials independently and in context), we considered the feedback provided by the children as very active, even though no researcher or adult was present during the time they were working with the probes. This is because the children actively engage with and elaborate on a tool or prototype to provide their feedback, for instance, by making video recordings. However, we clustered the methods in Figure 9 based on the descriptions of the authors of the reviewed papers and not on our own understanding or interpretation based on other literature. Accordingly, Figure 9 is based on empirical research rather than a methodological perspective, as was the case for Tsvyatkova and Storni (2019). The children's roles underlying this model (user, tester, informant, design partner) should not be understood as absolute, clear cut distinctions, but rather orient the reader regarding on which side of the continuum a certain method could be placed. Therefore, we present the different children's roles as circles, similar to the onion model introduced by Druin (2002). In that model, as one moves along the continuum, roles represent a lower level of involvement. Thus, a tester can perform the roles of both tester and user, while a design partner can perform all roles (Druin, 2002).

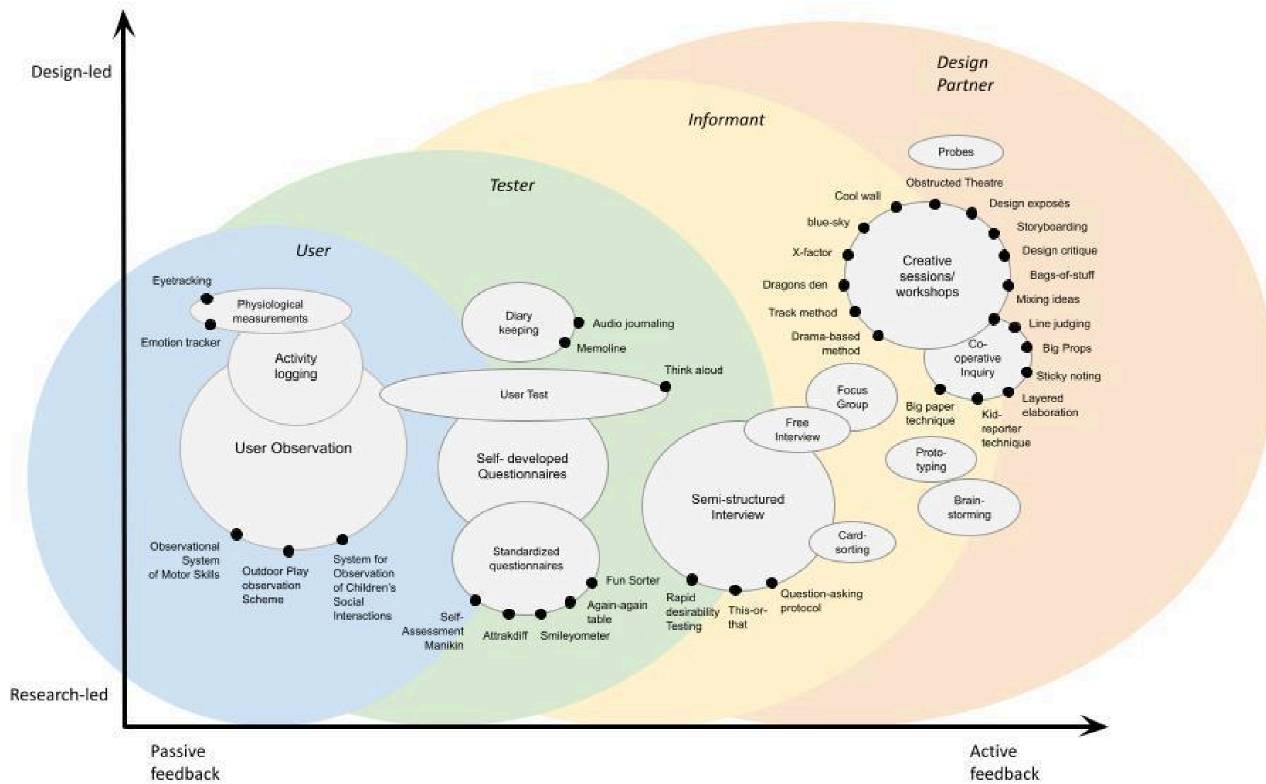


Figure 9. Illustration of an integrated understanding of interaction design methods for children.

5.7. Adult roles

Adults assumed different roles in the studies, which we coded based on the classification by Börjesson et al. (2015):

- *User:* a relative (e.g., sibling or parent) can also be the user of the technology developed.
- *Proxy*: people (e.g., parent, caregiver, teacher) surrounding the child are speaking for them.
- *Expert*: a professional (e.g., therapist, psychologist, teacher) aids in the design by speaking for a group of children. The difference to a proxy is that they do not seek to reflect a specific child's experiences.
- *Facilitator*: adults present during design activities to introduce the children and researchers or designers to each other, help build rapport, and provide practical assistance during the activity.

In addition to the roles described by Börjesson et al. (2015), we also coded whether the parents or legal guardians supported the children by e.g. escorting them to the testing location and waiting in a separate room. We called this role *support.* In the papers, mixed roles were most frequent (22.8%) (indicating any combination of roles, such as a parent acting as support while also help to facilitate the research), followed by the role of facilitator (12.1%), expert (10.7%), user (9.9%), proxy (3.3%) and support (1.5%). In 28 papers, we could not identify an adult role. Moreover, we found a relationship between different adult roles and the children's developmental group. We can see in Figure 10a that the role of expert is dominant for children with dyslexia, visual impairment, and Down Syndrome. Other developmental groups, such as diabetes or Autism Spectrum Disorder, have a wider spectrum of adult roles. In the mixed and typically developing groups, we see the widest variety of roles applied. In addition, Figure 10b shows that some adult roles are linked to the type of feedback children provide. The role of facilitator is often taken in methods where children have to provide active feedback. Adults take the role of expert more often in combination with verbal or passive feedback from children. Adults are least involved when children are asked to provide written feedback. In general, we observed a strong tendency towards mixed adult roles across all types of feedback.

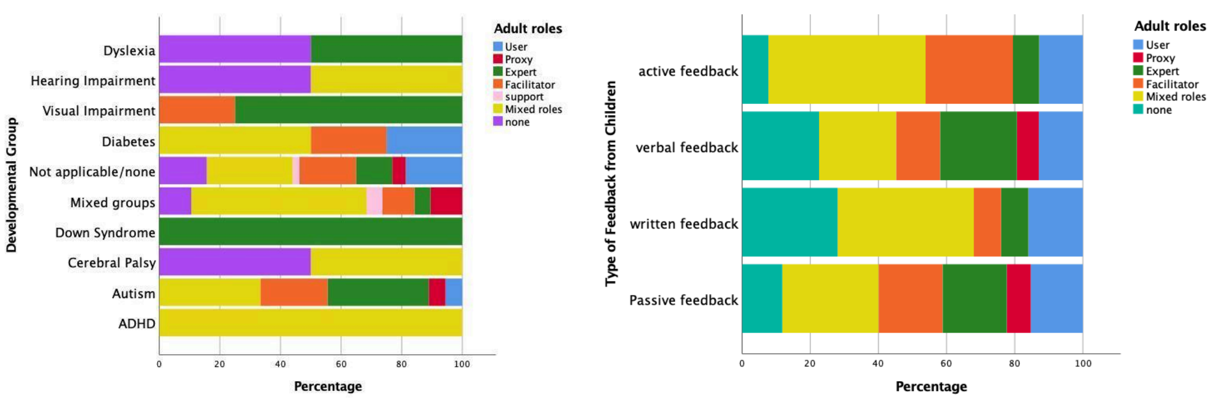


Figure 10a. Relation between children's developmental group and adult roles

Figure 10b. Relation between type of feedback gathered from children and adult roles

Furthermore, we were interested in the correlation between children's age and the roles adults take in the different papers. A point-biserial correlation was calculated to determine the relationship between the children's median age and adult roles. There was a positive correlation between age and adult role, which was statistically significant (r_{pb} = .138, n = 272, p = .028). Three outliers are included in the figure, from left to right, representing papers numbered 190, 27, and 241). We did not exclude those three papers with a mean age above 18 years from further analysis, as argued in more detail in our exclusion criteria in Section 3.2 . In the very active adult role as proxy (e.g. speaking for a child), the median children's age is significantly lower compared to the other adult roles (see Figure 11).

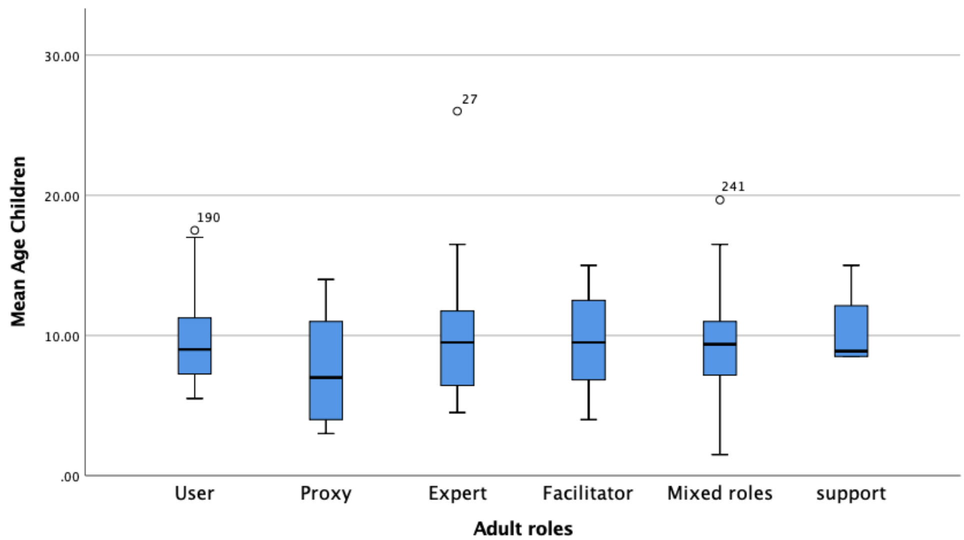


Figure 11. Boxplot representing the relation between adults' roles and children's age.

5.8. Ethics in research with children

Informed consent was reported in 42.7% of the papers. 19.5% state having received consent from parents only, 3.3% from children only, and 19.9% from both children and parents. In 56.6% of papers, no information or unclear information was provided about informed consent. Moreover, we counted the number of papers that reported ethical approval. Only 21% did report ethical approval, 1.8% clearly stated that they had not received ethical approval, and for 76.5% of reports, it was not clear whether ethical approval had been received or not.

6. Discussion

This systematic review of interaction design methods for children is based on a sample of 272 full empirical papers published in predominant venues in this field. We aimed to provide a comprehensive overview of methods used in the CCI field by surveying methods addressing the complete age range of childhood and adolescence (0-18 years), including developmentally diverse children, addressing stakeholders' roles and looking at methods and data triangulation patterns. In this section, we discuss future research challenges, which could serve as a guide for upcoming research in the CCI field.

6.1. Evolution of CCI methods

We identify an increasing number of CCI papers over the past 15 years, with conference proceedings from IDC and CHI the predominant publication venues. In terms of research locations, we found a strong tendency for natural settings, especially in the educational context (e.g., schools, preschools), compared to laboratory environments. As already noted in 2004 by Jensen and Skov (2005), the preference for natural settings prevails due to the usefulness and necessity of examining a product solution in a real-world environment. However, the lack of a controlled setting in the real world has led some researchers to prefer a controlled laboratory environment (Kjeldskov & Graham, 2003). The preference for natural settings is not surprising considering the types of products that are most commonly studied according to our review, namely interactive games and mobile applications. Such products lend themselves to investigation in the field, which can be essential in order to learn more about their usefulness and usability in realistic conditions. Similar to product categories for the adult population, where the most frequently studied products are mobile phone/applications and interactive games, newer technologies such as connected services/IoT and VR/AR have attracted increased interest over the past couple of years (Pettersson et al., 2018). However, these new technologies have not yet been used to support the evaluation of products or systems, which might be a future usage scenario. Furthermore, approximately half of the studies conducted in a natural setting evaluated a product or system at one moment in time (cross-sectional studies), compared to 41.9% that evaluated it several times (longitudinal studies). Petterson et al. (2018) observed a positive trend since 2010 for the adult population in this regard, so we might expect similar perspectives for the child population.

In terms of method triangulation, we observed a trend towards the use of more methods within the same study, as more than three-quarters of papers applied at least two methods and 34% at least three methods. Similar results were found for the adult population, with 72% of all papers using a triangulation approach (Petterson et al., 2018). The majority of papers in our study used a mixed-methods approach (qualitative and quantitative), which is not surprising when considering the arguments for applying triangulation, i.a. addressing the same problem from different angles to derive a fuller understanding of the problem and ultimately strengthen the results (Börjesson et al., 2015), or to "gain deeper insights". Petterson et al. (2018) mentioned yet another reason for triangulation, which is to better understand the results of other applied methods, e.g., using post-use interviews to make sense of observations. However, looking at the specific triangulation patterns outside the mixed-methods approaches, we saw that traditional and more research-led methods are often applied together, whereas we did not see the same trend for design-led methods. The focus seems to be on the conventional end with the research-led methods on one side of the coin and the design-led on the other. This might be due to resource constraints, or that design processes tend to be presented in a linear fashion, from a selected design method to a resulting design. There seems to be some uncertainty in the field as to what could be suitable combinations outside of traditional methods. Of the surveyed methods, user observation, interviews and questionnaires have the longest history, tracing back to early work in the humanities, which might explain why they are applied together most often (Pettersson et al., 2018).

As a new and interdisciplinary field, CCI has contributed to the development of participatory design methods, which are found in the design-led area with active user involvement (Sanders, 2008). The framework by Sanders (2008) illustrates this melting pot of approaches and methods. However, an open question concerns which methods qualify for triangulation and why. We therefore suggest that researchers go beyond data triangulation in general, with its obvious advantages (such as in terms of reliability) and instead look for new triangulation patterns along the two intersecting dimensions of design vs. research-led approaches and different levels of user involvement (from passive to active) (see Figure 9). An example could be to combine cooperative inquiry (design-led, active feedback) with standardized questionnaires (research-led, passive/written feedback). Combining research-led and design-led methods could, for instance, help design-led inquiry avoid confirmation bias and premature commitment, similar to the advantages of a mixed-methods approach (Pettersson et al., 2018). Furthermore, since design is contextual by nature, any attempt to generalize lessons learnt would benefit from a complementary research-led inquiry. This would allow us to extend our understanding by looking at the same problem from different perspectives, providing further inspiration for the development of technologies and products. However, exploring new method combinations remains a challenge, as many researchers might favor what has worked from them in the past.

6.2. Perspectives for diverse developmental groups

This review is the first to examine CCI methodology addressing the whole child population from 0 to 18 years, including developmentally diverse children. Similar to other reviews (Tsvyatkova & Storni, 2019; Yarosh et al., 2011), we see that not all age groups are evenly involved in the research and design process. Children between 7 and 13 years old (see Figure 1) are predominantly represented, while younger and older age groups have received less attention. This age focus seems independent of developmental group, as Yarosh's (Yarosh et al., 2011) findings indicate that typically developing children are generally between 6 and 12 years old, while developmentally diverse children involved in design are also within this age range (Tsvyatkova & Storni, 2019). We believe that by including children of different ages and from diverse developmental groups, we can provide a broader and more realistic picture of the current state of the art within the CCI field and can make explicit methodological advances that cut across developmental groups.

As such, we investigated sequential triangulation across different user groups, observing that qualitative methods are more often applied than quantitative methods for most user groups (see Figure 8). Similarly to other studies (Börjesson et al., 2015), we found that research methods with ASD children are more varied. In addition, research with ASD children applies methods from the participatory design field, a common approach and umbrella term for many creative techniques in which children are actively involved in the design process (Nesset & Large, 2004). Although this seems challenging for children with ASD, new methods are continuously being developed to involve them in the research and design process, such as the recently developed design critique tool (Frauenberger et al., 2013). In contrast, we observed that methods involving children with cerebral palsy more often involve gathering indirect feedback from methods such as activity logging or user testing (Saturno et al., 2015), which is also in line with Börjesson et al. (2015), who found user testing to be a dominant method along with prototyping and post-experience interviews. In general, we also see a mixed-methods trend within all groups of children, with visual impairment the only exception, which might be due to the low number of studies found for that group.

The most variation in method application was found in the typically developing group, which again might be due to the higher number of papers analysed for this group (75.3%). Given the predominance of typically developing children in recent research, we believe that the field should strive for greater inclusion when consolidating its methods. Most real-world technologies are not solely designed for one particular user group, and many school systems group developmentally diverse children together in inclusive classrooms. So why are we not developing methods that suit the needs of all children? Some authors state that different groups of children might require different approaches, such as providing a clear structure in the sessions, and different levels of active participation by caregivers, teachers and therapists based on different groups of children's unique needs (Börjesson et al., 2015). Linking different methods to particular skill sets children should possess, such as writing or expressing themselves verbally, could help with the selection of a suitable method. Another fruitful approach might be to investigate the environment in which children reside instead of focusing solely on their disability (Börjesson et al., 2015). We believe the first step in this direction can be taken by formulating appropriate alternatives to current methodologies by defining the rationale underlying each option and the specific user and usage description of context characteristics. We hope that the selection of appropriate methods will profit from this approach and create awareness of the benefits of employing heterogeneous methods. We do not suggest a *one-size-fits-all approach,* but wish to stimulate discussion about *whether* and *how* our methods are applicable for diverse populations of children, with the overarching goal of informing researchers of how a particular method can be used in different contexts, with different groups of children and stakeholders. Consequently, we could think about how the methods proposed for a specific research group could apply to different populations, whether developments and evidence have been replicated for different groups, and if the methods that are most commonly adapted to serve specific groups share certain characteristics.

6.3. Children's and adults' roles

It seems that work with some developmentally diverse groups does not follow the generally observed preference for active user involvement and instead relies on more classical approaches with a lower level of user involvement (e.g., indirect and written feedback) (Börjesson et al., 2015). Scholars researching the different roles taken by typically developing children in more detail have found that children are actively engaged in the ideation phase of design, whereas during evaluation, they are mainly considered as final users (Landoni et al., 2016). For developmentally diverse children, previous reviews have indicated that such children have been involved in a range of activities in several phases of the design process (Börjesson et al., 2015), which might indicate an active involvement overall. However, several papers also mentioned the difficulty of involving developmentally diverse children as design partners (Keay-Bright, 2007; Menzies, 2011). We found a relation between adults' roles and the types of feedback children provided. The facilitator role, representing the lowest degree of intervention by an adult, seems the most dominant across all types of feedback, but had the strongest relation with active feedback by the children. In contrast, more intervention by adults, for instance as an expert, is associated with the more passive involvement of children. As stated earlier, the methods that aim to place less of a burden upon children by studying them passively tend to have a greater research focus. In this context, it is also interesting to observe that the children's age has an influence on the roles adults take in the design process. However, the question remains open at this stage as to whether adults adopt certain roles based on the methods selected, or whether certain methods are selected because they are associated with certain adult roles. Overall, we see a strong potential benefit from studies that move away from the expert mindset in the direction of a participatory mindset, as we have seen an increasing trend of methodological developments in that area. It may be erroneously assumed that moving away from an expert mindset reduces the research focus. At this stage, we cannot answer the question of why this trend appears and why we are not seeing a research-led focus combined with a participatory mindset, as defined in the design research map by Sanders (2008). However, we would like to encourage future researchers to rethink the roles of adults throughout the design process and more actively make use of the wide spectrum of roles available. We also hope that researchers work to create a better balance of power relations by experimenting with new methods in the area of research-led and participatory mindset that would give children a stronger voice.

6.4. Reporting on the code of ethics

The salience of ethical issues when working with vulnerable groups such as children is not a new challenge. Consequently, we are rather concerned about this review's finding that more than half of papers did not report on informed consent and 76.5% did not report if they had received ethical approval to conduct the study. This finding is similar to a recently conducted systematic review of the top four conferences in the field of Human-Computer Interaction in the years 2018 and 2019, including the Conference on Human Factors in Computing Systems (CHI) (Van Mechelen et al., 2020). For CHI, 24.6% of papers mentioned an IRB or ethics committee review , similar to the 21% we found for our reviewed papers (Pater et al., 2021). Applying and reporting ethical procedures should not be a pragmatic choice, nor should we blindly assume that researchers must have fulfilled their duty. We as researchers have a responsibility to answer certain fundamental questions concerning the risks and burden to participating children. Some of the main conference venues for publishing interaction design work with children, such as the ACM Interaction Design and Children Conference, have formulated a strict requirement to include a special note on ethics and children in all research papers from 2018 onwards, following the adoption of theEuropean Union's General Data Protection Regulation (GDPR) (Calder et al., 2017). In this note, paper authors should state how the participating children were selected, what consent processes were followed (e.g., did they consent and if so, what they were told), how they were treated, how data sharing was communicated, and any additional ethical considerations.

The GDPR has also contributed to shaping the ACM Code of Ethics, as Principle 1.6 “Respect privacy” now includes a call for a more general requirement for informed consent procedures (Gotterbarn et al., 2017). According to the ACM, informed consent procedures should go beyond ensuring that users understand what data is being collected and used; they should actually give them the ability to consent or withhold consent from data collection. Informed consent should ideally be a proactive agreement with the user about the type, content, and usage of data collected about them (Gotterbarn et al., 2017). Van Mechelen et al. (Van Mechelen et al., 2014) developed CHECk, a value checklist created for use prior to and at the start of design activities to help CCI researchers critically consider their values when involving children in design projects and examine how best to explain participatory design activities to children to aid informed consent. These developments are promising; however, we strongly encourage CCI researchers to pay sufficient attention to ethical issues when working with children. The successful application of a code of ethics should reflect the values of our community in a way that can help researchers make appropriate ethical decisions.

7. Limitations

This review had several limitations. First, when searching for relevant publications according to the defined search query, we narrowed our scope down to two main databases (ACM and Scopus). Our analysis might thus have missed relevant papers, such as papers conducted within disciplinary backgrounds not typically referenced in ACM and Scopus. We also required the presence of “child” or related keywords (as presented in Table 1) in the title, which might exclude some papers from specialized venues such as IDC, which do not require children to be mentioned in the title. We also only cover published literature in the English language, as the language of international research. Community-specific approaches published in other languages might also have been overlooked. Second, due to the ambiguity of what is considered by the authors themselves as a tool, a method and a technique, we did not distinguish between those terms in our analysis. Third, even though the first two authors worked together to double-code a number of papers, from which we calculated the interrater reliability, and the third and fourth authors did an independant cross-check on a sample of papers, most papers were single-coded by the first author. This might have introduced some bias. However, discussing uncertainties and cross-checking the coding scheme with independent researchers in the field helped to further reduce bias in the interpretation of some codes. Finally, the interpretation of results related to diverse developmental groups should be taken with caution, as some of those groups were represented in a limited number of papers. For instance, only one study involving children with ADHD was included in the review.

8. Conclusion

In this paper, we conducted a structured literature review to examine the state of the art in interaction design methods for children. Our final corpus consisted of 272 empirical papers. Based on our analysis, we built a systematic understanding of recent empirical methodologies in the CCI field. Our results indicate a preference for conducting field studies, mainly focusing on product categories such as interactive games and professional software. Furthermore, we observed an increase of studies focusing on newer technologies such as VR/AR in recent years and identified a preference for evaluating products repeatedly over time, with more than half of researchers applying a combination of two or more methods to design or evaluate their product. However, as we found that combining various research-led methods, such as user observation and semi-structured interviews, was most common, we would like to challenge future research to look for new combinations of methods across the wide continuum of research- and design-led approaches by, for instance, mixing different roles of stakeholder involvement. The challenges reported in former reviews published in the field also emphasizes the need for more methods and activities suitable for children of younger ages and diverse abilities, which is also in line with our findings. A strong effort to develop more methods in the participatory design area, such as creative sessions with children, is also clearly visible, especially when looking at the development of new tools and techniques. It seems desirable for this trend to continue into the future and be extended to all developmental groups. Along with this, we would like to draw attention to the need to establish best research practices by reporting ethical procedures when involving children in the research and design process. With this in mind, we look forward to future trends and challenges in the application of CCI methods towards a truly child-centred approach.

9. Acknowledgments

The authors would like to thank the HCI research group from the University of Luxembourg, which provided extensive feedback on the coding scheme. Also, we gratefully acknowledge the support of the

Leibniz ScienceCampus Bremen Digital Public Health (lsc-diph.de), which is jointly funded by the Leibniz Association (W4/2018), the Federal State of Bremen and the Leibniz Institute for Prevention Research and Epidemiology – BIPS.

10. References Chapter 2

Alexandros, L., & Michalis, X. (2013). The physiological measurements as a critical indicator in users' experience evaluation. *Proceedings of the 17th Panhellenic Conference on Informatics - PCI '13*, 258. https://doi.org/10.1145/2491845.2491883

Alhumaidan, H., Lo, K. P. Y., & Selby, A. (2018). Co-designing with children a collaborative augmented reality book based on a primary school textbook. *International Journal of Child-Computer Interaction*, *15*, 24–36. https://doi.org/10.1016/j.ijcci.2017.11.005

Al-Samarraie, H., & Hurmuzan, S. (2018). A review of brainstorming techniques in higher education. *Thinking Skills and Creativity*, *27*, 78–91. https://doi.org/10.1016/j.tsc.2017.12.002

Antle, A. N., McLaren, E.-S., Fiedler, H., & Johnson, N. (2019). Evaluating the Impact of a Mobile Neurofeedback App for Young Children at School and Home. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–13. https://doi.org/10.1145/3290605.3300266

Atterer, R., Wnuk, M., & Schmidt, A. (2006). Knowing the user's every move: User activity tracking for website usability evaluation and implicit interaction. *Proceedings of the 15th International Conference on World Wide Web - WWW '06*, 203. https://doi.org/10.1145/1135777.1135811

Bastien, J. M. C. (2010). Usability testing: A review of some methodological and technical aspects of the method. *International Journal of Medical Informatics*, *79*(4), e18–e23. https://doi.org/10.1016/j.ijmedinf.2008.12.004

Bateman, A., Zhao, O. K., Bajcsy, A. V., Jennings, M. C., Toth, B. N., Cohen, A. J., Horton, E. L., Khattar, A., Kuo, R. S., Lee, F. A., Lim, M. K., Migasiuk, L. W., Renganathan, R., Zhang, A., & Oliveira, M. A. (2018). A user-centered design and analysis of an electrostatic haptic touchscreen system for students with visual impairments. *International Journal of Human-Computer Studies*, *109*, 102–111. https://doi.org/10.1016/j.ijhcs.2017.09.004

Blandford, A., Furniss, D., & Makri, S. (2016). *Qualitative HCI research: Going behind the scenes*. Morgan&Claypool.

Börjesson, P., Barendregt, W., Eriksson, E., & Torgersson, O. (2015). Designing technology for and with developmentally diverse children: A systematic literature review. *Proceedings of the 14th International Conference on Interaction Design and Children - IDC '15*, 79–88. https://doi.org/10.1145/2771839.2771848

Bowen, S., Sustar, H., Wolstenholme, D., & Dearden, A. (2013). Engaging teenagers productively in service design. *International Journal of Child-Computer Interaction*, *1*(3–4), 71–81. https://doi.org/10.1016/j.ijcci.2014.02.001

Boynton, P. M., & Greenhalgh, T. (2004). Selecting, designing, and developing your questionnaire. *BMJ*, *328*(7451), 1312–1315. https://doi.org/10.1136/bmj.328.7451.1312

Brule, E., Bailly, G., Brock, A., Valentin, F., Denis, G., & Jouffrais, C. (2016). MapSense: Multi-Sensory Interactive Maps for Children Living with Visual Impairments. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 445–457. https://doi.org/10.1145/2858036.2858375

Burns, T., & Gottschalk, F. (2019). *What do we know about children and technology?* OECD. file:///Users/florencekristin.lehn/Desktop/PhD/Literature/Books/OECD-what%20we%20know%20about%20children%20and%20technology.pdf

Calder, A., IT Governance (Organization), Privacy Team, European Union, & European Parliament. (2017). *EU general data protection regulation (GDPR): An implementation and compliance guide*. https://www.itgovernanceusa.com/download/EU-GDPR-Implementation-and-Compliance-Guide.pdf

Cerezo, R., Calderón, V., & Romero, C. (2019). A holographic mobile-based application for practicing pronunciation of basic English vocabulary for Spanish speaking children. *International Journal of Human-Computer Studies*, *124*, 13–25. https://doi.org/10.1016/j.ijhcs.2018.11.009

Chaparro, B., & Hinkle, V. (2008). Card-Sorting: What You Need to Know about Analyzing and Interpreting Card Sorting Results. *Usability News*, *10*(2).

Chu, S. L., Quek, F., Bhangaonkar, S., Ging, A. B., & Sridharamurthy, K. (2015). Making the Maker: A Means-to-an-Ends approach to nurturing the Maker mindset in elementary-aged children. *International Journal of Child-Computer Interaction*, *5*, 11–19. https://doi.org/10.1016/j.ijcci.2015.08.002

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed). Sage Publications.

Desjardins, A., & Wakkary, R. (2011). How children represent sustainability in the home. *Proceedings of the 10th International Conference on Interaction Design and Children - IDC '11*, 37–45. https://doi.org/10.1145/1999030.1999035

Druin, A. (Ed.). (1999). *The design of children's technology*. Morgan Kaufmann Publishers.

Druin, A. (2002). The Role of Children in the Design of New Technology. *Behaviour and Information Technology*, *21*(1), 1–25.

Fails, J. A., Guha, M. L., & Druin, A. (2012). Methods and Techniques for Involving Children in the Design of New Technology for Children. *Foundations and Trends in Human–Computer Interaction*, *6*(2), 85–166. https://doi.org/10.1561/1100000018

Fails, J. A., Pera, M. S., Anuyah, O., Kennington, C., Wright, K. L., & Bigirimana, W. (2019). Query Formulation Assistance for Kids: What is Available, When to Help & What Kids Want. *Proceedings of the Interaction Design and Children on ZZZ - IDC '19*, 109–120. https://doi.org/10.1145/3311927.3323131

Fleck, S., Baraudon, C., Frey, J., Lainé, T., & Hachet, M. (2018). "Teegi's so cute!": Assessing the pedagogical potential of an interactive tangible interface for schoolchildren. *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 143–156. https://doi.org/10.1145/3202185.3202731

Fletcher-Watson, S., Pain, H., Hammond, S., Humphry, A., & McConachie, H. (2016). Designing for young children with autism spectrum disorder: A case study of an iPad app. *International Journal of Child-Computer Interaction*, *7*, 1–14. https://doi.org/10.1016/j.ijcci.2016.03.002

Follmer, S., & Ishii, H. (2012). KidCAD: Digitally remixing toys through tangible tools. *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems - CHI '12*, 2401. https://doi.org/10.1145/2207676.2208403

Fransen, S., & Markopoulos, P. (2010). Let robots do the talking. *Proceedings of the 9th International Conference on Interaction Design and Children - IDC '10*, 59. https://doi.org/10.1145/1810543.1810551

Frauenberger, C., Good, J., Alcorn, A., & Pain, H. (2013). Conversing through and about technologies: Design critique as an opportunity to engage children with autism and broaden research(er) perspectives. *International Journal of Child-Computer Interaction*, *1*(2), 38–49. https://doi.org/10.1016/j.ijcci.2013.02.001

Frauenberger, C., Makhaeva, J., & Spiel, K. (2016). Designing Smart Objects with Autistic Children: Four Design Exposès. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 130–139. https://doi.org/10.1145/2858036.2858050

Frauenberger, C., Makhaeva, J., & Spiel, K. (2017). Blending Methods: Developing Participatory Design Sessions for Autistic Children. *Proceedings of the 2017 Conference on Interaction Design and Children*, 39–49. https://doi.org/10.1145/3078072.3079727

Garzotto, F. (2008). Broadening children’s involvement as design partners: From technology to “experience.” *Proceedings of the 7th International Conference on Interaction Design and Children - IDC ’08*, 186. https://doi.org/10.1145/1463689.1463755

Gotterbarn, D., Bruckman, A., Flick, C., Miller, K., & Wolf, M. J. (2017). ACM code of ethics: A guide for positive action. *Communications of the ACM*, *61*(1), 121–128. https://doi.org/10.1145/3173016

Gray, S., Robertson, J., & Rajendran, G. (2015). BrainQuest: An active smart phone game to enhance executive function. *Proceedings of the 14th International Conference on Interaction Design and Children - IDC ’15*, 59–68. https://doi.org/10.1145/2771839.2771846

Hailpern, J., Karahalios, K., & Halle, J. (2009). Creating a spoken impact: Encouraging vocalization through audio visual feedback in children with ASD. *Proceedings of the 27th International Conference on Human Factors in Computing Systems - CHI 09*, 453. https://doi.org/10.1145/1518701.1518774

Hamidi, F., Saenz, K., & Baljko, M. (2014). Sparkles of brilliance: Incorporating cultural and social context in codesign of digital artworks. *Proceedings of the 2014 Conference on Interaction Design and Children - IDC ’14*, 77–84. https://doi.org/10.1145/2593968.2593977

Harms, K. J., Cosgrove, D., Gray, S., & Kelleher, C. (2013). Automatically generating tutorials to enable middle school children to learn programming independently. *Proceedings of the 12th International Conference on Interaction Design and Children - IDC ’13*, 11–19. https://doi.org/10.1145/2485760.2485764

Hiniker, A., Lee, B., Sobel, K., & Choe, E. K. (2017). Plan & Play: Supporting Intentional Media Use in Early Childhood. *Proceedings of the 2017 Conference on Interaction Design and Children*, 85–95. https://doi.org/10.1145/3078072.3079752

Hourcade, J. P. (2015). *Child-Computer Interaction*.

Hudson, N., Lafreniere, B., Chilana, P. K., & Grossman, T. (2018). Investigating How Online Help and Learning Resources Support Children’s Use of 3D Design Software. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI ’18*, 1–14. https://doi.org/10.1145/3173574.3173831

Jensen, J. J., & Skov, M. B. (2005). A review of research methods in children’s technology design. *Proceeding of the 2005 Conference on Interaction Design and Children - IDC ’05*, 80–87. https://doi.org/10.1145/1109540.1109551

Jones, S., Hall, L., & Hilton, J. (2012). Exploring children’s “indexical encounter” with real and digitised archive photographs using tablet and large flat screen technologies. *Proceedings of the 11th*

International Conference on Interaction Design and Children - IDC '12, 30. https://doi.org/10.1145/2307096.2307101

Kawas, S., Yuan, Y., DeWitt, A., Jin, Q., Kirchner, S., Bilger, A., Grantham, E., Kientz, J. A., Tartaro, A., & Yarosh, S. (2020). Another decade of IDC research: Examining and reflecting on values and ethics. *Proceedings of the Interaction Design and Children Conference*, 205–215. https://doi.org/10.1145/3392063.3394436

Keay-Bright, W. E. (2007). The Reactive Colours Project: Demonstrating Participatory and Collaborative Design Methods for the Creation of Software for Autistic Children. *Design Principles and Practices: An International Journal—Annual Review*, *1*(2), 7–16. https://doi.org/10.18848/1833-1874/CGP/v01i02/37623

Kjeldskov, J., & Graham, C. (2003). A Review of Mobile HCI Research Methods. *Proceedings of the 5th International Conference on Mobile Human-Computer Interaction (MobileHCI03)*, *2795*, 317–335. https://doi.org/10.1007/978-3-540-45233-1_23

Kory-Westlund, J. M., & Breazeal, C. (2019). Assessing Children's Perceptions and Acceptance of a Social Robot. *Proceedings of the Interaction Design and Children on ZZZ - IDC '19*, 38–50. https://doi.org/10.1145/3311927.3323143

Kumar, P., Vitak, J., Chetty, M., Clegg, T. L., Yang, J., McNally, B., & Bonsignore, E. (2018). Co-designing online privacy-related games and stories with children. *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 67–79. https://doi.org/10.1145/3202185.3202735

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159. https://doi.org/10.2307/2529310

Landoni, M., Rubegni, E., Nicol, E., & Read, J. (2016). How Many Roles Can Children Play? *Proceedings of the The 15th International Conference on Interaction Design and Children - IDC '16*, 720–725. https://doi.org/10.1145/2930674.2932222

Landry, P., Minsky, J., Castañer, M., Camerino, O., Rodriguez-Arregui, R., Ormo, E., & Pares, N. (2013). Design strategy to stimulate a diversity of motor skills for an exergame addressed to children. *Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13*, 84–91. https://doi.org/10.1145/2485760.2485781

Leduc-Mills, B., & Eisenberg, M. (2011). The UCube: A child-friendly device for introductory three-dimensional design. *Proceedings of the 10th International Conference on Interaction Design and Children - IDC '11*, 72–80. https://doi.org/10.1145/1999030.1999039

Leite, I., Pereira, A., & Lehman, J. F. (2017). Persistent Memory in Repeated Child-Robot Conversations. *Proceedings of the 2017 Conference on Interaction Design and Children*, 238–247. https://doi.org/10.1145/3078072.3079728

Lim, Y.-K., Stolterman, E., & Tenenberg, J. (2008). The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas. *ACM Transactions on Computer-Human Interaction*, *15*(2), 1–27. https://doi.org/10.1145/1375761.1375762

Liszio, S., & Masuch, M. (2017). Virtual Reality MRI: Playful Reduction of Children's Anxiety in MRI Exams. *Proceedings of the 2017 Conference on Interaction Design and Children*, 127–136. https://doi.org/10.1145/3078072.3079713

Ma, Y., Bekker, T., Ren, X., Hu, J., & Vos, S. (2018). Effects of playful audio augmentation on teenagers' motivations in cooperative physical play. *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 43–54. https://doi.org/10.1145/3202185.3202729

Maqsood, S., Mekhail, C., & Chiasson, S. (2018). A day in the life of jos: A web-based game to increase children's digital literacy. *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 241–252. https://doi.org/10.1145/3202185.3202753

Markopoulos, P., Read, J. C., MacFarlane, S., & Hoysniemi, J. (2008). *Evaluating Children's Interactive Products*. Morgan Kaufmann.

Martin-Niedecken, A. L. (2018). Designing for bodily interplay: Engaging with the adaptive social exertion game "plunder planet." *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 19–30. https://doi.org/10.1145/3202185.3202740

Mazzone, E., Iivari, N., Tikkanen, R., Read, J. C., & Beale, R. (2010). Considering context, content, management, and engagement in design activities with children. *Proceedings of the 9th International Conference on Interaction Design and Children - IDC '10*, 108. https://doi.org/10.1145/1810543.1810556

McRoberts, S., Yuan, Y., Watson, K., & Yarosh, S. (2019). Behind the Scenes: Design, Collaboration, and Video Creation with Youth. *Proceedings of the Interaction Design and Children on ZZZ - IDC '19*, 173–184. https://doi.org/10.1145/3311927.3323134

Menzies, R. (2011). Developing for autism with user-centred design. *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '11*, 313. https://doi.org/10.1145/2049536.2049624

Metatla, O., Maggioni, E., Cullen, C., & Obrist, M. (2019). "Like Popcorn": Crossmodal Correspondences Between Scents, 3D Shapes and Emotions in Children. *Proceedings of the 2019 CHI Conference* *on Human Factors in Computing Systems - CHI ’19*, 1–13. https://doi.org/10.1145/3290605.3300689

Michaelis, J. E., & Mutlu, B. (2019). Supporting Interest in Science Learning with a Social Robot. *Proceedings of the Interaction Design and Children on ZZZ - IDC ’19*, 71–82. https://doi.org/10.1145/3311927.3323154

Miles, M. B., Hubermann, A. M., & Saldana, J. (1994). Qualitative data analysis: A methods sourcebook (3rd ed.). Sage.

Moher, D., Liberati, A., Tetzlaff, J., & Altmann, D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Plos Medicine*, *6*(7), 1–6. https://doi.org/10.1371/ journal.pmed.1000097

MoraGuiard, J., & Pares, N. (2014). “Child as the measure of all things”: The body as a referent in designing a museum exhibit to understand the nanoscale. *Proceedings of the 2014 Conference on Interaction Design and Children - IDC ’14*, 27–36. https://doi.org/10.1145/2593968.2593985

Nacher, V., Cáliz, D., Jaen, J., & Martínez, L. (2018). Examining the Usability of Touch Screen Gestures for Children With Down Syndrome. *Interacting with Computers*, *30*(3), 258–272. https://doi.org/10.1093/iwc/iwy011

Nesset, V., & Large, A. (2004). Children in the information technology design process: A review of theories and their applications. *Library & Information Science Research*, *26*(2), 140–161. https://doi.org/10.1016/j.lisr.2003.12.002

Norouzi, Kinnula, & Iivari. (2019). Interaction Order and Historical Body Shaping Children’s Making Projects—A Literature Review. *Multimodal Technologies and Interaction*, *3*(4), 71. https://doi.org/10.3390/mti3040071

Pater, J., Coupe, A., Pfafman, R., Phelan, C., Toscos, T., & Jacobs, M. (2021). Standardizing Reporting of Participant Compensation in HCI: A Systematic Literature Review and Recommendations for the Field. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–16. https://doi.org/10.1145/3411764.3445734

Pettersson, I., Lachner, F., Frison, A.-K., Riener, A., & Butz, A. (2018). A Bermuda Triangle?: A Review of Method Application and Triangulation in User Experience Evaluation. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI ’18*, 1–16. https://doi.org/10.1145/3173574.3174035

Raffle, H., Spasojevic, M., Ballagas, R., Revelle, G., Horii, H., Follmer, S., Go, J., Reardon, E., Mori, K., & Kaye, J. (2010). Family story play: Reading with young children (and elmo) over a distance.

Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10, 1583. https://doi.org/10.1145/1753326.1753563

Read, J. C., Fitton, D., & Hortton, M. (2014). Giving ideas an equal chance: Inclusion and representation in participatory design with children. *Proceedings of the 2014 Conference on Interaction Design and Children - IDC '14*, 105–114. https://doi.org/10.1145/2593968.2593986

Read, J., Gregory, P., MacFarlane, S., McManus, B., Gray, P., & Patel, R. (2002). An Investigation of Participatory Design with Children – Informant, Balanced and Facilitated Design. *Procs Interaction Design and Children*, 53–64.

Rick, J., Marshall, P., & Yuill, N. (2011). Beyond one-size-fits-all: How interactive tabletops support collaborative learning. *Proceedings of the 10th International Conference on Interaction Design and Children - IDC '11*, 109–117. https://doi.org/10.1145/1999030.1999043

Roto, V., Law, E., Vermeeren, A., & Hoonhout, J. (Eds.). (2011). *The UX White Paper. Outcome of the Dagstuhl Seminar on Demarcating User Experience*. https://drops.dagstuhl.de/opus/volltexte/2011/2949/pdf/10373_AbstractsCollection.2949.pdf

Sanders, L. (2008). ON MODELINGAn evolving map of design practice and design research. *Interactions*, *15*(6), 13–17. https://doi.org/10.1145/1409040.1409043

Saturno, C. E., Ramirez, A. R. G., Conte, M. J., Farhat, M., & Piucco, E. C. (2015). An augmentative and alternative communication tool for children and adolescents with cerebral palsy. *Behaviour & Information Technology*, *34*(6), 632–645. https://doi.org/10.1080/0144929X.2015.1019567

Schaper, M.-M., Santos, M., & Pares, N. (2018). Orchestrating experts' assumptions and children's values in the design of Virtual Heritage experiences. *International Journal of Child-Computer Interaction*, *17*, 5–15. https://doi.org/10.1016/j.ijcci.2018.02.001

Scharf, F., Winkler, T., Hahn, C., Wolters, C., & Herczeg, M. (2012). Tangicons 3.0: An educational non-competitive collaborative game. *Proceedings of the 11th International Conference on Interaction Design and Children - IDC '12*, 144. https://doi.org/10.1145/2307096.2307113

Segura, E. M., Cramer, H., Gomes, P. F., Nylander, S., & Paiva, A. (2012). Revive!: Reactions to migration between different embodiments when playing with robotic pets. *Proceedings of the 11th International Conference on Interaction Design and Children - IDC '12*, 88. https://doi.org/10.1145/2307096.2307107

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 - IDC '12*, 70. https://doi.org/10.1145/2307096.2307105

Sim, G., Nouwen, M., Vissers, J., Horton, M., Slegers, K., & Zaman, B. (2016). Using the MemoLine to capture changes in user experience over time with children. *International Journal of Child-Computer Interaction*, *8*, 1–14. https://doi.org/10.1016/j.ijcci.2016.07.001

Sonne, T., Müller, J., Marshall, P., Obel, C., & Grønbæk, K. (2016). Changing Family Practices with Assistive Technology: MOBERO Improves Morning and Bedtime Routines for Children with ADHD. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 152–164. https://doi.org/10.1145/2858036.2858157

Soute, I., Lagerström, S., & Markopoulos, P. (2013). Rapid prototyping of outdoor games for children in an iterative design process. *Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13*, 74–83. https://doi.org/10.1145/2485760.2485779

Spiel, K., Frauenberger, C., Keyes, O., & Fitzpatrick, G. (2019). Agency of Autistic Children in Technology Research—A Critical Literature Review. *ACM Transactions on Computer-Human Interaction*, *26*(6), 1–40. https://doi.org/10.1145/3344919

Spiel, K., & Gerling, K. (2021). The Purpose of Play: How HCI Games Research Fails Neurodivergent Populations. *ACM Transactions on Computer-Human Interaction*, *28*(2), 1–40. https://doi.org/10.1145/3432245

Thoring, K., Luippold, C., & Mueller, R. (2013). Opening the Cultural Probes Box: A Critical Reflection and Analysis of the Cultural Probes Method. *In Proceedings of the 5th International Congress of International Association of Societies of Design Research*, 222–233.

Tsvyatkova, D., & Storni, C. (2019). A review of selected methods, techniques and tools in Child–Computer Interaction (CCI) developed/adapted to support children's involvement in technology development. *International Journal of Child-Computer Interaction*, *22*, 100148. https://doi.org/10.1016/j.ijcci.2019.100148

Vacca, R. (2019). Brokering Data: Co-Designing Technology with Latina Teens to Support Communication with Parents: Leveraging Cultural Practices of Latinx Youth through Co-Design. *Proceedings of the Interaction Design and Children on ZZZ - IDC '19*, 197–207. https://doi.org/10.1145/3311927.3323142

Van Mechelen, M., Baykal, G. E., Dindler, C., & Eriksson, E. (2020). 18 Years of ethics in child-computer interaction research: A systematic literature review. *Proceedings of the Interaction Design and Children Conference*, 161–183.

Van Mechelen, M., Sim, G., Zaman, B., Gregory, P., Slegers, K., & Horton, M. (2014). Applying the CHECk tool to participatory design sessions with children. *Proceedings of the 2014 Conference on Interaction Design and Children - IDC '14*, 253–256. https://doi.org/10.1145/2593968.2610465

Van Mechelen, M., Zaman, B., Laenen, A., & Abeele, V. V. (2015). Challenging group dynamics in participatory design with children: Lessons from social interdependence theory. *Proceedings of the 14th International Conference on Interaction Design and Children - IDC '15*, 219–228. https://doi.org/10.1145/2771839.2771862

Ventä-Olkkonen, L., Hartikainen, H., Norouzi, B., Iivari, N., & Kinnula, M. (2019). A Literature Review of the Practice of Educating Children about Technology Making. *IFIP Conference on Human-Computer Interaction*, 418–441.

Woodward, J., McFadden, Z., Shiver, N., Ben-hayon, A., Yip, J. C., & Anthony, L. (2018). Using Co-Design to Examine How Children Conceptualize Intelligent Interfaces. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–14. https://doi.org/10.1145/3173574.3174149

Wyeth, P., & MacColl, I. (2010). Noising around: Investigations in mobile learning. *Proceedings of the 9th International Conference on Interaction Design and Children - IDC '10*, 147. https://doi.org/10.1145/1810543.1810560

Yarosh, S., Radu, I., Hunter, S., & Rosenbaum, E. (2011). Examining values: An analysis of nine years of IDC research. *Proceedings of the 10th International Conference on Interaction Design and Children - IDC '11*, 136–144. https://doi.org/10.1145/1999030.1999046

Yip, J. C., Clegg, T., Ahn, J., Uchidiuno, J. O., Bonsignore, E., Beck, A., Pauw, D., & Mills, K. (2016). The Evolution of Engagements and Social Bonds During Child-Parent Co-design. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 3607–3619. https://doi.org/10.1145/2858036.2858380

Yip, J. C., Sobel, K., Gao, X., Hishikawa, A. M., Lim, A., Meng, L., Ofiana, R. F., Park, J., & Hiniker, A. (2019). Laughing is Scary, but Farting is Cute: A Conceptual Model of Children's Perspectives of Creepy Technologies. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–15. https://doi.org/10.1145/3290605.3300303

Zaman, B., & Abeele, V. V. (2010). Laddering with young children in User eXperience evaluations: Theoretical groundings and a practical case. *Proceedings of the 9th International Conference on Interaction Design and Children - IDC '10*, 156. https://doi.org/10.1145/1810543.1810561

Zhou, Y., Percival, G., Wang, X., Wang, Y., & Zhao, S. (2011). MOGCLASS: Evaluation of a collaborative system of mobile devices for classroom music education of young children. *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI '11*, 523. https://doi.org/10.1145/1978942.1979016

11. Appendix: Selected Review Papers

1. Abrahamson, Dor, and Dragan Trninic. “Toward an Embodied-Interaction Design Framework for Mathematical Concepts.” In *Proceedings of the 10th International Conference on Interaction Design and Children - IDC ’11*, 1–10. Ann Arbor, Michigan: ACM Press, 2011. https://doi.org/10.1145/1999030.1999031.
2. Alcorn, Alyssa M., Helen Pain, and Judith Good. “Discrepancies in a Virtual Learning Environment: Something ‘Worth Communicating about’ for Young Children with ASC?” In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC ’13*, 56–65. New York, New York: ACM Press, 2013. https://doi.org/10.1145/2485760.2485783.
3. Alessandrini, Andrea, Alessandro Cappelletti, and Massimo Zancanaro. “Audio-Augmented Paper for Therapy and Educational Intervention for Children with Autistic Spectrum Disorder.” *International Journal of Human-Computer Studies* 72, no. 4 (April 2014): 422–30. https://doi.org/10.1016/j.ijhcs.2013.12.001.
4. Alhumaidan, Haifa, Kathy Pui Ying Lo, and Andrew Selby. “Co-Designing with Children a Collaborative Augmented Reality Book Based on a Primary School Textbook.” *International Journal of Child-Computer Interaction* 15 (March 2018): 24–36. https://doi.org/10.1016/j.ijcci.2017.11.005.
5. Almanji, Amur, T. Claire Davies, and N. Susan Stott. “Using Cursor Measures to Investigate the Effects of Impairment Severity on Cursor Control for Youths with Cerebral Palsy.” *International Journal of Human-Computer Studies* 72, no. 3 (March 2014): 349–57. https://doi.org/10.1016/j.ijhcs.2013.12.003.
6. Anthony, Lisa, Brown, Quincy, Nias, Jaye, Tate, Berthel. “Children (and Adults) Benefit from Visual Feedback during Gesture Interaction on Mobile Touchscreen Devices.” *International Journal of Child-Computer Interaction* 6 (December 2015): 17–27. https://doi.org/10.1016/j.ijcci.2016.01.002.
7. Anthony, Lisa, Brown, Quincy, Nias, Jaye, Tate, Berthel. “Examining the Need for Visual Feedback during Gesture Interaction on Mobile Touchscreen Devices for Kids.” In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC*

’13, 157–64. New York, New York: ACM Press, 2013. https://doi.org/10.1145/2485760.2485775.

8. Antle, Alissa N., Leslie Chesick, Aaron Levisohn, Srilekha Kirshnamachari Sridharan, and Perry Tan. “Using Neurofeedback to Teach Self-Regulation to Children Living in Poverty.” In *Proceedings of the 14th International Conference on Interaction Design and Children - IDC ’15*, 119–28. Boston, Massachusetts: ACM Press, 2015. https://doi.org/10.1145/2771839.2771852.
9. Antle, Alissa N., Leslie Chesick, Srilekha Kirshnamachari Sridharan, and Emily Cramer. “East Meets West: A Mobile Brain-Computer System That Helps Children Living in Poverty Learn to Self-Regulate.” *Personal and Ubiquitous Computing* 22, no. 4 (August 2018): 839–66. https://doi.org/10.1007/s00779-018-1166-x.
10. Antle, Alissa N., Elgin-Skye McLaren, Holly Fiedler, and Naomi Johnson. “Evaluating the Impact of a Mobile Neurofeedback App for Young Children at School and Home.” In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI ’19*, 1–13. Glasgow, Scotland Uk: ACM Press, 2019. https://doi.org/10.1145/3290605.3300266.
11. Antle, Alissa N., Jillian L. Warren, Aaron May, Min Fan, and Alyssa F. Wise. “Emergent Dialogue: Eliciting Values during Children’s Collaboration with a Tabletop Game for Change.” In *Proceedings of the 2014 Conference on Interaction Design and Children - IDC ’14*, 37–46. Aarhus, Denmark: ACM Press, 2014. https://doi.org/10.1145/2593968.2593971.
12. Antle, Alissa N., Alyssa F. Wise, and Kristine Nielsen. “Towards Utopia: Designing Tangibles for Learning.” In *Proceedings of the 10th International Conference on Interaction Design and Children - IDC ’11*, 11–20. Ann Arbor, Michigan: ACM Press, 2011. https://doi.org/10.1145/1999030.1999032.
13. Apostolellis, Panagiotis, and Doug A. Bowman. “Small Group Learning with Games in Museums: Effects of Interactivity as Mediated by Cultural Differences.” In *Proceedings of the 14th International Conference on Interaction Design and Children - IDC ’15*, 160–69. Boston, Massachusetts: ACM Press, 2015. https://doi.org/10.1145/2771839.2771856.
14. Apostolellis, Panagiotis, Marjee Chmiel, and Doug A. Bowman. “‘Pump That Press!’: Design Evaluation of Audience Interaction Using Collaborative Digital and Physical Games.” In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 31–42. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202736.

15. Arteaga, Sonia M., Mo Kudeki, Adrienne Woodworth, and Sri Kurniawan. "Mobile System to Motivate Teenagers' Physical Activity." In *Proceedings of the 9th International Conference on Interaction Design and Children - IDC '10*, 1. Barcelona, Spain: ACM Press, 2010. https://doi.org/10.1145/1810543.1810545.
16. Asselborn, Thibault, Arzu Guneysu, Khalil Mrini, Elmira Yadollahi, Ayberk Ozgur, Wafa Johal, and Pierre Dillenbourg. "Bringing Letters to Life: Handwriting with Haptic-Enabled Tangible Robots." In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 219–30. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202747.
17. Avontuur, Tetske, Rian de Jong, Eveline Brink, Yves Florack, Iris Soute, and Panos Markopoulos. "Play It Our Way: Customization of Game Rules in Children's Interactive Outdoor Games." In *Proceedings of the 2014 Conference on Interaction Design and Children - IDC '14*, 95–104. Aarhus, Denmark: ACM Press, 2014. https://doi.org/10.1145/2593968.2593973.
18. Axelrod, Daryl B., and Jennifer Kahn. "Intergenerational Family Storytelling and Modeling with Large-Scale Data Sets." In *Proceedings of the Interaction Design and Children on ZZZ - IDC '19*, 352–60. Boise, ID, USA: ACM Press, 2019. https://doi.org/10.1145/3311927.3323153.
19. Badillo-Urquiola, Karla, Diva Smriti, Brenna McNally, Evan Golub, Elizabeth Bonsignore, and Pamela J. Wisniewski. "Stranger Danger!: Social Media App Features Co-Designed with Children to Keep Them Safe Online." In *Proceedings of the Interaction Design and Children on ZZZ - IDC '19*, 394–406. Boise, ID, USA: ACM Press, 2019. https://doi.org/10.1145/3311927.3323133.
20. Bakker, Saskia, Alissa N. Antle, and Elise van den Hoven. "Identifying Embodied Metaphors in Children's Sound-Action Mappings." In *Proceedings of the 8th International Conference on Interaction Design and Children - IDC '09*, 140. Como, Italy: ACM Press, 2009. https://doi.org/10.1145/1551788.1551812.
21. Baranauskas, M. Cecília C., and Julian E. Gutiérrez Posada. "Tangible and Shared Storytelling: Searching for the Social Dimension of Constructionism." In *Proceedings of the 2017 Conference on Interaction Design and Children*, 193–203. Stanford California USA: ACM, 2017. https://doi.org/10.1145/3078072.3079743.
22. Bartoli, Laura, Clara Corradi, Franca Garzotto, and Matteo Valoriani. "Exploring Motion-Based Touchless Games for Autistic Children's Learning." In *Proceedings of the 12th*

International Conference on Interaction Design and Children - IDC ’13, 102–11. New York, New York: ACM Press, 2013. https://doi.org/10.1145/2485760.2485774.

23. Bartoli, Laura, Franca Garzotto, Mirko Gelsomini, Luigi Oliveto, and Matteo Valoriani. “Designing and Evaluating Touchless Playful Interaction for ASD Children.” In *Proceedings of the 2014 Conference on Interaction Design and Children - IDC ’14*, 17–26. Aarhus, Denmark: ACM Press, 2014. https://doi.org/10.1145/2593968.2593976.
24. Bateman, Amelia, Oliver K. Zhao, Andrea V. Bajcsy, Mathew C. Jennings, Bryan N. Toth, Alexa J. Cohen, Emily L. Horton, et al. “A User-Centered Design and Analysis of an Electrostatic Haptic Touchscreen System for Students with Visual Impairments.” *International Journal of Human-Computer Studies* 109 (January 2018): 102–11. https://doi.org/10.1016/j.ijhcs.2017.09.004.
25. Beheshti, Elham, Katya Borgos-Rodriguez, and Anne Marie Piper. “Supporting Parent-Child Collaborative Learning through Haptic Feedback Displays.” In *Proceedings of the Interaction Design and Children on ZZZ - IDC ’19*, 58–70. Boise, ID, USA: ACM Press, 2019. https://doi.org/10.1145/3311927.3323137.
26. Benton, Laura, Hilary Johnson, Emma Ashwin, Mark Brosnan, and Beate Grawemeyer. “Developing IDEAS: Supporting Children with Autism within a Participatory Design Team.” In *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems - CHI ’12*, 2599. Austin, Texas, USA: ACM Press, 2012. https://doi.org/10.1145/2207676.2208650.
27. Benton, Laura, George Varotsis, and Asimina Vasalou. “Leading by Example: Exploring the Influence of Design Examples on Children’s Creative Ideation.” *International Journal of Human-Computer Studies* 122 (February 2019): 174–83. https://doi.org/10.1016/j.ijhcs.2018.09.007.
28. Benton, Laura, Asimina Vasalou, Rilla Khaled, Hilary Johnson, and Daniel Gooch. “Diversity for Design: A Framework for Involving Neurodiverse Children in the Technology Design Process.” In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI ’14*, 3747–56. Toronto, Ontario, Canada: ACM Press, 2014. https://doi.org/10.1145/2556288.2557244.
29. Benveniste, Samuel, Pierre Jouvelot, Edith Lecourt, and Renaud Michel. “Designing Wiimprovisation for Mediation in Group Music Therapy with Children Suffering from Behavioral Disorders.” In *Proceedings of the 8th International Conference on Interaction*

Design and Children - IDC '09, 18. Como, Italy: ACM Press, 2009. https://doi.org/10.1145/1551788.1551793.

30. Beşevli, Ceylan, Elif Salman, Tilbe Goksun, Hakan Urey, and Oğuzhan Özcan. "MaR-T: Designing a Projection-Based Mixed Reality System for Nonsymbolic Math Development of Preschoolers: Guided by Theories of Cognition and Learning." In *Proceedings of the Interaction Design and Children on ZZZ - IDC '19*, 280–92. Boise, ID, USA: ACM Press, 2019. https://doi.org/10.1145/3311927.3323147.
31. Bhattacharya, Arpita, Mirko Gelsomini, Patricia Pérez-Fuster, Gregory D. Abowd, and Agata Rozga. "Designing Motion-Based Activities to Engage Students with Autism in Classroom Settings." In *Proceedings of the 14th International Conference on Interaction Design and Children - IDC '15*, 69–78. Boston, Massachusetts: ACM Press, 2015. https://doi.org/10.1145/2771839.2771847.
32. Bodén, Marie, Andrew Dekker, Stephen Viller, and Ben Matthews. "Augmenting Play and Learning in the Primary Classroom." In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13*, 228–36. New York, New York: ACM Press, 2013. https://doi.org/10.1145/2485760.2485767.
33. Bonsignore, Elizabeth, Derek Hansen, Kari Kraus, Amanda Visconti, June Ahn, and Allison Druin. "Playing for Real: Designing Alternate Reality Games for Teenagers in Learning Contexts." In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13*, 237–46. New York, New York: ACM Press, 2013. https://doi.org/10.1145/2485760.2485788.
34. Bowen, Simon, Helena Sustar, Daniel Wolstenholme, and Andy Dearden. "Engaging Teenagers Productively in Service Design." *International Journal of Child-Computer Interaction* 1, no. 3–4 (September 2013): 71–81. https://doi.org/10.1016/j.ijcci.2014.02.001.
35. Brule, Emeline, Gilles Bailly, Anke Brock, Frederic Valentin, Grégoire Denis, and Christophe Jouffrais. "MapSense: Multi-Sensory Interactive Maps for Children Living with Visual Impairments." In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 445–57. San Jose California USA: ACM, 2016. https://doi.org/10.1145/2858036.2858375.
36. Cabrera, Lautaro, John H. Maloney, and David Weintrop. "Programs in the Palm of Your Hand: How Live Programming Shapes Children's Interactions with Physical Computing

Devices." In *Proceedings of the Interaction Design and Children on ZZZ - IDC '19*, 227–36. Boise, ID, USA: ACM Press, 2019. https://doi.org/10.1145/3311927.3323138.

37. Fleck, Stéphanie, Charlotte Baraudon, Jérémy Frey, Thibault Lainé, and Martin Hachet. "'Teegi's so Cute!': Assessing the Pedagogical Potential of an Interactive Tangible Interface for Schoolchildren." In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 143–56. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202731.

38. McCarthy, Betsy, Linlin Li, Michelle Tiu, and Sara Atienza. "PBS KIDS Mathematics Transmedia Suites in Preschool Homes." In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13*, 128–36. New York, New York: ACM Press, 2013. https://doi.org/10.1145/2485760.2485777.

39. Roffo, Giorgio, Dong-Bach Vo, Mohammad Tayarani, Maki Rooksby, Alessandra Sorrentino, Simona Di Folco, Helen Minnis, Stephen Brewster, and Alessandro Vinciarelli. "Automating the Administration and Analysis of Psychiatric Tests: The Case of Attachment in School Age Children." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–12. Glasgow, Scotland Uk: ACM Press, 2019. https://doi.org/10.1145/3290605.3300825.

40. Soler-Adillon, Joan, Jaume Ferrer, and Narcís Parés. "A Novel Approach to Interactive Playgrounds: The Interactive Slide Project." In *Proceedings of the 8th International Conference on Interaction Design and Children - IDC '09*, 131. Como, Italy: ACM Press, 2009. https://doi.org/10.1145/1551788.1551811.

41. Cahill, Clara, Alex Kuhn, Shannon Schmoll, Wan-Tzu Lo, Brenna McNally, and Chris Quintana. "Mobile Learning in Museums: How Mobile Supports for Learning Influence Student Behavior." In *Proceedings of the 10th International Conference on Interaction Design and Children - IDC '11*, 21–28. Ann Arbor, Michigan: ACM Press, 2011. https://doi.org/10.1145/1999030.1999033.

42. Carter, Elizabeth J., Jennifer Hyde, and Jessica K. Hodgins. "Investigating the Effects of Interactive Features for Preschool Television Programming." In *Proceedings of the 2017 Conference on Interaction Design and Children*, 97–106. Stanford California USA: ACM, 2017. https://doi.org/10.1145/3078072.3079717.

43. Castañer, Marta, Oleguer Camerino, Pascal Landry, and Narcis Pares. "Quality of Physical Activity of Children in Exergames: Sequential Body Movement Analysis and Its Implications

for Interaction Design." *International Journal of Human-Computer Studies* 96 (December 2016): 67–78. https://doi.org/10.1016/j.ijhcs.2016.07.007.

44. Celis, Véronique, Jelle Husson, Vero Vanden Abeele, Leen Loyez, Lieven Van den Audenaeren, Pol Ghesquière, Ann Goeleven, Jan Wouters, and Luc Geurts. "Translating Preschoolers' Game Experiences into Design Guidelines via a Laddering Study." In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13*, 147–56. New York, New York: ACM Press, 2013. https://doi.org/10.1145/2485760.2485772.

45. Cerezo, Rebeca, Vicente Calderón, and Cristóbal Romero. "A Holographic Mobile-Based Application for Practicing Pronunciation of Basic English Vocabulary for Spanish Speaking Children." *International Journal of Human-Computer Studies* 124 (April 2019): 13–25. https://doi.org/10.1016/j.ijhcs.2018.11.009.

46. Chandra, Shruti, Pierre Dillenbourg, and Ana Paiva. "Classification of Children's Handwriting Errors for the Design of an Educational Co-Writer Robotic Peer." In *Proceedings of the 2017 Conference on Interaction Design and Children*, 215–25. Stanford California USA: ACM, 2017. https://doi.org/10.1145/3078072.3079750.

47. Chase, Kiera, and Dor Abrahamson. "Reverse Scaffolding: A Constructivist Design Architecture for Mathematics Learning with Educational Technology." In *Proceedings of the 14th International Conference on Interaction Design and Children - IDC '15*, 189–98. Boston, Massachusetts: ACM Press, 2015. https://doi.org/10.1145/2771839.2771859.

48. Cheng, Yi, Kate Yen, Yeqi Chen, Sijin Chen, and Alexis Hiniker. "Why Doesn't It Work?: Voice-Driven Interfaces and Young Children's Communication Repair Strategies." In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 337–48. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202749.

49. Chipman, Gene, Jerry Alan Fails, Allison Druin, and Mona Leigh Guha. "Paper vs. Tablet Computers: A Comparative Study Using Tangible Flags." In *Proceedings of the 10th International Conference on Interaction Design and Children - IDC '11*, 29–36. Ann Arbor, Michigan: ACM Press, 2011. https://doi.org/10.1145/1999030.1999034.

50. Christensen, Peter Knøsgaard, Christoffer Øland Skovgaard, and Marianne Graves Petersen. "Together Together: Combining Shared and Separate Activities in Designing Technology for Family Life." In *Proceedings of the Interaction Design and Children on ZZZ - IDC '19*, 374–85. Boise, ID, USA: ACM Press, 2019. https://doi.org/10.1145/3311927.3323141.

51. Chu, Sharon Lynn, Brittany Garcia, and Beth Nam. "Understanding Context in Children's Use of Smartwatches for Everyday Science Reflections." In *Proceedings of the Interaction Design and Children on ZZZ - IDC '19*, 83–93. Boise, ID, USA: ACM Press, 2019. https://doi.org/10.1145/3311927.3323144.

52. Chu, Sharon Lynn, Francis Quek, Sourabh Bhangaonkar, Amy Boettcher Ging, and Kumar Sridharamurthy. "Making the Maker: A Means-to-an-Ends Approach to Nurturing the Maker Mindset in Elementary-Aged Children." *International Journal of Child-Computer Interaction* 5 (September 2015): 11–19. https://doi.org/10.1016/j.ijcci.2015.08.002.

53. Cibrian, Franceli L., Monica Tentori, and Ana I. Martínez-García. "Hunting Relics: A Persuasive Exergame to Promote Collective Exercise in Young Children." *International Journal of Human-Computer Interaction* 32, no. 3 (March 3, 2016): 277–94. https://doi.org/10.1080/10447318.2016.1136180.

54. Clegg, Tamara, Elizabeth Bonsignore, Jason Yip, Helene Gelderblom, Alex Kuhn, Tobin Valenstein, Becky Lewittes, and Allison Druin. "Technology for Promoting Scientific Practice and Personal Meaning in Life-Relevant Learning." In *Proceedings of the 11th International Conference on Interaction Design and Children - IDC '12*, 152. Bremen, Germany: ACM Press, 2012. https://doi.org/10.1145/2307096.2307114.

55. Coenraad, Merijke, Jen Palmer, Diana Franklin, and David Weintrop. "Enacting Identities: Participatory Design as a Context for Youth to Reflect, Project, and Apply Their Emerging Identities." In *Proceedings of the Interaction Design and Children on ZZZ - IDC '19*, 185–96. Boise, ID, USA: ACM Press, 2019. https://doi.org/10.1145/3311927.3323148.

56. Colombo, Luca, and Monica Landoni. "A Diary Study of Children's User Experience with EBooks Using Flow Theory as Framework." In *Proceedings of the 2014 Conference on Interaction Design and Children - IDC '14*, 135–44. Aarhus, Denmark: ACM Press, 2014. https://doi.org/10.1145/2593968.2593978.

57. Coyle, David, Nicola McGlade, Gavin Doherty, and Gary O'Reilly. "Exploratory Evaluations of a Computer Game Supporting Cognitive Behavioural Therapy for Adolescents." In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI '11*, 2937. Vancouver, BC, Canada: ACM Press, 2011. https://doi.org/10.1145/1978942.1979378.

58. Cramer, Meg, Russell Beauregard, and Mayank Sharma. "An Investigation of Purpose Built Netbooks for Primary School Education." In *Proceedings of the 8th International Conference*

on Interaction Design and Children - IDC '09, 36. Como, Italy: ACM Press, 2009. https://doi.org/10.1145/1551788.1551796.

59. Crescenzi Lanna, Lucrezia, and Mariona Grané Oro. "Touch Gesture Performed by Children under 3 Years Old When Drawing and Coloring on a Tablet." *International Journal of Human-Computer Studies* 124 (April 2019): 1–12. https://doi.org/10.1016/j.ijhcs.2018.11.008.
60. Cullen, Clare, and Oussama Metatla. "Co-Designing Inclusive Multisensory Story Mapping with Children with Mixed Visual Abilities." In *Proceedings of the Interaction Design and Children on ZZZ - IDC '19*, 361–73. Boise, ID, USA: ACM Press, 2019. https://doi.org/10.1145/3311927.3323146.
61. D'Angelo, Sarah, D. Harmon Pollock, and Michael Horn. "Fishing with Friends: Using Tabletop Games to Raise Environmental Awareness in Aquariums." In *Proceedings of the 14th International Conference on Interaction Design and Children - IDC '15*, 29–38. Boston, Massachusetts: ACM Press, 2015. https://doi.org/10.1145/2771839.2771843.
62. Dasgupta, Sayamindu. "From Surveys to Collaborative Art: Enabling Children to Program with Online Data." In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13*, 28–35. New York, New York: ACM Press, 2013. https://doi.org/10.1145/2485760.2485784.
63. Davis, Richard Lee, Melisa Orta Martinez, Oliver Schneider, Karon E. MacLean, Allison M. Okamura, and Paulo Blikstein. "The Haptic Bridge: Towards a Theory for Haptic-Supported Learning." In *Proceedings of the 2017 Conference on Interaction Design and Children*, 51–60. Stanford California USA: ACM, 2017. https://doi.org/10.1145/3078072.3079755.
64. Desai, Shital, Alethea Blackler, and Vesna Popovic. "Children's Embodied Intuitive Interaction — Design Aspects of Embodiment." *International Journal of Child-Computer Interaction* 21 (September 2019): 89–103. https://doi.org/10.1016/j.ijcci.2019.06.001.
65. Desjardins, Audrey, and Ron Wakkary. "How Children Represent Sustainability in the Home." In *Proceedings of the 10th International Conference on Interaction Design and Children - IDC '11*, 37–45. Ann Arbor, Michigan: ACM Press, 2011. https://doi.org/10.1145/1999030.1999035.
66. Donker, Afke, and Pieter Reitsma. "Drag-and-Drop Errors in Young Children's Use of the Mouse." *Interacting with Computers* 19, no. 2 (March 2007): 257–66. https://doi.org/10.1016/j.intcom.2006.05.008.

67. Druga, Stefania, Randi Williams, Hae Won Park, and Cynthia Breazeal. "How Smart Are the Smart Toys?: Children and Parents' Agent Interaction and Intelligence Attribution." In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 231–40. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202741.

68. Druin, Allison, Elizabeth Foss, Leshell Hatley, Evan Golub, Mona Leigh Guha, Jerry Fails, and Hilary Hutchinson. "How Children Search the Internet with Keyword Interfaces." In *Proceedings of the 8th International Conference on Interaction Design and Children - IDC '09*, 89. Como, Italy: ACM Press, 2009. https://doi.org/10.1145/1551788.1551804.

69. Duarte, Ana Maria Bustamante, Nina Brendel, Auriol Degbelo, and Christian Kray. "Participatory Design and Participatory Research: An HCI Case Study with Young Forced Migrants." *ACM Transactions on Computer-Human Interaction* 25, no. 1 (February 24, 2018): 1–39. https://doi.org/10.1145/3145472.

70. Duveskog, Marcus, Kati Kemppainen, Roman Bednarik, and Erkki Sutinen. "Designing a Story-Based Platform for HIV and AIDS Counseling with Tanzanian Children." In *Proceedings of the 8th International Conference on Interaction Design and Children - IDC '09*, 27. Como, Italy: ACM Press, 2009. https://doi.org/10.1145/1551788.1551794.

71. Eriksson, Eva, Gökçe Elif Baykal, Staffan Björk, and Olof Torgersson. "Using Gameplay Design Patterns with Children in the Redesign of a Collaborative Co-Located Game." In *Proceedings of the Interaction Design and Children on ZZZ - IDC '19*, 15–25. Boise, ID, USA: ACM Press, 2019. https://doi.org/10.1145/3311927.3323155.

72. Ervasti, Mari, Juhani Laitakari, and Mika Hillukkala. "'I Want to Know Where My Child Is at All Times' – Field Study of a Location-Aware Safety Service for Schoolchildren." *Behaviour & Information Technology* 35, no. 10 (October 2, 2016): 833–52. https://doi.org/10.1080/0144929X.2016.1201144.

73. Escobedo, Lizbeth, Catalina Ibarra, Jehu Hernandez, Mariana Alvelais, and Monica Tentori. "Smart Objects to Support the Discrimination Training of Children with Autism." *Personal and Ubiquitous Computing* 18, no. 6 (August 2014): 1485–97. https://doi.org/10.1007/s00779-013-0750-3.

74. Escobedo, Lizbeth, David H. Nguyen, LouAnne Boyd, Sen Hirano, Alejandro Rangel, Daniel Garcia-Rosas, Monica Tentori, and Gillian Hayes. "MOSOCO: A Mobile Assistive Tool to Support Children with Autism Practicing Social Skills in Real-Life Situations." In *Proceedings*

of the 2012 ACM Annual Conference on Human Factors in Computing Systems - CHI '12, 2589. Austin, Texas, USA: ACM Press, 2012. https://doi.org/10.1145/2207676.2208649.

75. Fails, Jerry Alan, Maria Soledad Pera, Oghenemaro Anuyah, Casey Kennington, Katherine Landau Wright, and William Bigirimana. "Query Formulation Assistance for Kids: What Is Available, When to Help & What Kids Want." In *Proceedings of the Interaction Design and Children on ZZZ - IDC '19*, 109–20. Boise, ID, USA: ACM Press, 2019. https://doi.org/10.1145/3311927.3323131.
76. Fan, Min, Alissa N. Antle, Maureen Hoskyn, Carman Neustaedter, and Emily S. Cramer. "Why Tangibility Matters: A Design Case Study of At-Risk Children Learning to Read and Spell." In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 1805–16. Denver Colorado USA: ACM, 2017. https://doi.org/10.1145/3025453.3026048.
77. Farr, William, Nicola Yuill, Eric Harris, and Steve Hinske. "In My Own Words: Configuration of Tangibles, Object Interaction and Children with Autism." In *Proceedings of the 9th International Conference on Interaction Design and Children - IDC '10*, 30. Barcelona, Spain: ACM Press, 2010. https://doi.org/10.1145/1810543.1810548.
78. Fernaeus, Ylva, Maria Håkansson, Mattias Jacobsson, and Sara Ljungblad. "How Do You Play with a Robotic Toy Animal?: A Long-Term Study of Pleo." In *Proceedings of the 9th International Conference on Interaction Design and Children - IDC '10*, 39. Barcelona, Spain: ACM Press, 2010. https://doi.org/10.1145/1810543.1810549.
79. Ferron, Michela, Chiara Leonardi, Paolo Massa, Gianluca Schiavo, Amy L. Murphy, and Elisabetta Farella. "A Walk on the Child Side: Investigating Parents' and Children's Experience and Perspective on Mobile Technology for Outdoor Child Independent Mobility." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–12. Glasgow, Scotland Uk: ACM Press, 2019. https://doi.org/10.1145/3290605.3300827.
80. Fitton, Dan, and Janet C. Read. "Creating a Framework to Support the Critical Consideration of Dark Design Aspects in Free-to-Play Apps." In *Proceedings of the Interaction Design and Children on ZZZ - IDC '19*, 407–18. Boise, ID, USA: ACM Press, 2019. https://doi.org/10.1145/3311927.3323136.
81. Fitton, Daniel, Janet C Read, Gavin Sim, and Brendan Cassidy. "Co-Designing Voice User Interfaces with Teenagers in the Context of Smart Homes." In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 55–66. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202744.

82. Flannery, Louise P., Brian Silverman, Elizabeth R. Kazakoff, Marina Umaschi Bers, Paula Bontá, and Mitchel Resnick. "Designing ScratchJr: Support for Early Childhood Learning through Computer Programming." In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13*, 1–10. New York, New York: ACM Press, 2013. https://doi.org/10.1145/2485760.2485785.

83. Fletcher-Watson, S., H. Pain, S. Hammond, A. Humphry, and H. McConachie. "Designing for Young Children with Autism Spectrum Disorder: A Case Study of an IPad App." *International Journal of Child-Computer Interaction* 7 (January 2016): 1–14. https://doi.org/10.1016/j.ijcci.2016.03.002.

84. Follmer, Sean, and Hiroshi Ishii. "KidCAD: Digitally Remixing Toys through Tangible Tools." In *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems - CHI '12*, 2401. Austin, Texas, USA: ACM Press, 2012. https://doi.org/10.1145/2207676.2208403.

85. Follmer, Sean, Hayes Raffle, Janet Go, Rafael Ballagas, and Hiroshi Ishii. "Video Play: Playful Interactions in Video Conferencing for Long-Distance Families with Young Children." In *Proceedings of the 9th International Conference on Interaction Design and Children - IDC '10*, 49. Barcelona, Spain: ACM Press, 2010. https://doi.org/10.1145/1810543.1810550.

86. Fransen, Sjef, and Panos Markopoulos. "Let Robots Do the Talking." In *Proceedings of the 9th International Conference on Interaction Design and Children - IDC '10*, 59. Barcelona, Spain: ACM Press, 2010. https://doi.org/10.1145/1810543.1810551.

87. Frauenberger, Christopher, Judith Good, Alyssa Alcorn, and Helen Pain. "Conversing through and about Technologies: Design Critique as an Opportunity to Engage Children with Autism and Broaden Research(Er) Perspectives." *International Journal of Child-Computer Interaction* 1, no. 2 (May 2013): 38–49. https://doi.org/10.1016/j.ijcci.2013.02.001.

88. Frauenberger, Christopher, Judith Good, Alyssa Alcorn, and Helen Pain. "Supporting the Design Contributions of Children with Autism Spectrum Conditions." In *Proceedings of the 11th International Conference on Interaction Design and Children - IDC '12*, 134. Bremen, Germany: ACM Press, 2012. https://doi.org/10.1145/2307096.2307112.

89. Frauenberger, Christopher, Julia Makhaeva, and Katta Spiel. "Blending Methods: Developing Participatory Design Sessions for Autistic Children." In *Proceedings of the 2017 Conference on Interaction Design and Children*, 39–49. Stanford California USA: ACM, 2017. https://doi.org/10.1145/3078072.3079727.

90. Frauenberger, Christopher, Makhaeva, Spiel, Katta. “Designing Smart Objects with Autistic Children: Four Design Exposès.” In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 130–39. San Jose California USA: ACM, 2016. https://doi.org/10.1145/2858036.2858050.

91. Frauenberger, Christopher, Katta Spiel, Laura Scheepmaker, and Irene Posch. “Nurturing Constructive Disagreement - Agonistic Design with Neurodiverse Children.” In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI ’19*, 1–11. Glasgow, Scotland Uk: ACM Press, 2019. https://doi.org/10.1145/3290605.3300501.

92. Freed, Natalie, Winslow Burleson, Hayes Raffle, Rafael Ballagas, and Naomi Newman. “User Interfaces for Tangible Characters: Can Children Connect Remotely through Toy Perspectives?” In *Proceedings of the 9th International Conference on Interaction Design and Children - IDC ’10*, 69. Barcelona, Spain: ACM Press, 2010. https://doi.org/10.1145/1810543.1810552.

93. Freier, Nathan G. “Children Attribute Moral Standing to a Personified Agent.” In *Proceeding of the Twenty-Sixth Annual CHI Conference on Human Factors in Computing Systems - CHI ’08*, 343. Florence, Italy: ACM Press, 2008. https://doi.org/10.1145/1357054.1357113.

94. Garzotto, Franca. “Broadening Children’s Involvement as Design Partners: From Technology to ‘Experience.’” In *Proceedings of the 7th International Conference on Interaction Design and Children - IDC ’08*, 186. Chicago, Illinois: ACM Press, 2008. https://doi.org/10.1145/1463689.1463755.

95. Garzotto, Franca, and Manuel Bordogna. “Paper-Based Multimedia Interaction as Learning Tool for Disabled Children.” In *Proceedings of the 9th International Conference on Interaction Design and Children - IDC ’10*, 79. Barcelona, Spain: ACM Press, 2010. https://doi.org/10.1145/1810543.1810553.

96. Gerling, Kathrin, Kieran Hicks, Michael Kalyn, Adam Evans, and Conor Linehan. “Designing Movement-Based Play With Young People Using Powered Wheelchairs.” In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4447–58. San Jose California USA: ACM, 2016. https://doi.org/10.1145/2858036.2858070.

97. Gerling, Kathrin M., Conor Linehan, Ben Kirman, Michael R. Kalyn, Adam B. Evans, and Kieran C. Hicks. “Creating Wheelchair-Controlled Video Games: Challenges and Opportunities When Involving Young People with Mobility Impairments and Game Design

Experts.” *International Journal of Human-Computer Studies* 94 (October 2016): 64–73. https://doi.org/10.1016/j.ijhcs.2015.08.009.

98. Giusti, Leonardo, Massimo Zancanaro, Eynat Gal, and Patrice L. (Tamar) Weiss. “Dimensions of Collaboration on a Tabletop Interface for Children with Autism Spectrum Disorder.” In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI ’11*, 3295. Vancouver, BC, Canada: ACM Press, 2011. https://doi.org/10.1145/1978942.1979431.

99. Golsteijn, Connie, and Elise van den Hoven. “Facilitating Parent-Teenager Communication through Interactive Photo Cubes.” *Personal and Ubiquitous Computing* 17, no. 2 (February 2013): 273–86. https://doi.org/10.1007/s00779-011-0487-9.

100. González-Ferreras, César, David Escudero-Mancebo, Mario Corrales-Astorgano, Lourdes Aguilar-Cuevas, and Valle Flores-Lucas. “Engaging Adolescents with Down Syndrome in an Educational Video Game.” *International Journal of Human–Computer Interaction* 33, no. 9 (September 2, 2017): 693–712. https://doi.org/10.1080/10447318.2017.1278895.

101. Gossen, Tatiana, Rene Müller, Sebastian Stober, and Andreas Nürnberger. “Search Result Visualization with Characters for Children.” In *Proceedings of the 2014 Conference on Interaction Design and Children - IDC ’14*, 125–34. Aarhus, Denmark: ACM Press, 2014. https://doi.org/10.1145/2593968.2593983.

102. Gourlet, Pauline. “Children’s Conversation with Experience: Making Emotional Imprints.” In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 313–24. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202734.

103. Gray, Stuart, Judy Robertson, and Gnanathusharan Rajendran. “BrainQuest: An Active Smart Phone Game to Enhance Executive Function.” In *Proceedings of the 14th International Conference on Interaction Design and Children - IDC ’15*, 59–68. Boston, Massachusetts: ACM Press, 2015. https://doi.org/10.1145/2771839.2771846.

104. Hailpern, Joshua, Karrie Karahalios, and James Halle. “Creating a Spoken Impact: Encouraging Vocalization through Audio Visual Feedback in Children with ASD.” In *Proceedings of the 27th International Conference on Human Factors in Computing Systems - CHI 09*, 453. Boston, MA, USA: ACM Press, 2009. https://doi.org/10.1145/1518701.1518774.

105. Hair, Adam, Penelope Monroe, Beena Ahmed, Kirrie J. Ballard, and Ricardo Gutierrez-Osuna. “Apraxia World: A Speech Therapy Game for Children with Speech Sound

Disorders.” In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 119–31. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202733.

106. Hamidi, Foad, Karla Saenz, and Melanie Baljko. “Sparkles of Brilliance: Incorporating Cultural and Social Context in Codesign of Digital Artworks.” In *Proceedings of the 2014 Conference on Interaction Design and Children - IDC ’14*, 77–84. Aarhus, Denmark: ACM Press, 2014. https://doi.org/10.1145/2593968.2593977.

107. Harms, Kyle J., Dennis Cosgrove, Shannon Gray, and Caitlin Kelleher. “Automatically Generating Tutorials to Enable Middle School Children to Learn Programming Independently.” In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC ’13*, 11–19. New York, New York: ACM Press, 2013. https://doi.org/10.1145/2485760.2485764.

108. Hashish, Yasmeen, Andrea Bunt, and James E. Young. “Involving Children in Content Control: A Collaborative and Education-Oriented Content Filtering Approach.” In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI ’14*, 1797–1806. Toronto, Ontario, Canada: ACM Press, 2014. https://doi.org/10.1145/2556288.2557128.

109. Hernandez, Hamilton A., T.C. Nicholas Graham, Darcy Fehlings, Lauren Switzer, Zi Ye, Quentin Bellay, Md Ameer Hamza, Cheryl Savery, and Tadeusz Stach. “Design of an Exergaming Station for Children with Cerebral Palsy.” In *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems - CHI ’12*, 2619. Austin, Texas, USA: ACM Press, 2012. https://doi.org/10.1145/2207676.2208652.

110. Hernandez, Hamilton A., Zi Ye, T.C. Nicholas Graham, Darcy Fehlings, and Lauren Switzer. “Designing Action-Based Exergames for Children with Cerebral Palsy.” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI ’13*, 1261. Paris, France: ACM Press, 2013. https://doi.org/10.1145/2470654.2466164.

111. Heslop, Philip, Ahmed Kharrufa, Madeline Balaam, David Leat, Paul Dolan, and Patrick Olivier. “Learning Extended Writing: Designing for Children’s Collaboration.” In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC ’13*, 36–45. New York, New York: ACM Press, 2013. https://doi.org/10.1145/2485760.2485765.

112. Hiniker, Alexis, Jon E. Froehlich, Mingrui Zhang, and Erin Beneteau. “Anchored Audio Sampling: A Seamless Method for Exploring Children’s Thoughts During Deployment Studies.” In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*

- *CHI ’19*, 1–13. Glasgow, Scotland Uk: ACM Press, 2019. https://doi.org/10.1145/3290605.3300238.

113. Hiniker, Alexis, Bongshin Lee, Kiley Sobel, and Eun Kyoung Choe. “Plan & Play: Supporting Intentional Media Use in Early Childhood.” In *Proceedings of the 2017 Conference on Interaction Design and Children*, 85–95. Stanford California USA: ACM, 2017. https://doi.org/10.1145/3078072.3079752.

114. Hiniker, Alexis, Kiley Sobel, Sungsoo Ray Hong, Hyewon Suh, India Irish, Daniella Kim, and Julie A. Kientz. “Touchscreen Prompts for Preschoolers: Designing Developmentally Appropriate Techniques for Teaching Young Children to Perform Gestures.” In *Proceedings of the 14th International Conference on Interaction Design and Children - IDC ’15*, 109–18. Boston, Massachusetts: ACM Press, 2015. https://doi.org/10.1145/2771839.2771851.

115. Hitron, Tom, Yoav Orlev, Iddo Wald, Ariel Shamir, Hadas Erel, and Oren Zuckerman. “Can Children Understand Machine Learning Concepts?: The Effect of Uncovering Black Boxes.” In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI ’19*, 1–11. Glasgow, Scotland Uk: ACM Press, 2019. https://doi.org/10.1145/3290605.3300645.

116. Hoffman, Guy, Ayelet Gal-Oz, Shlomi David, and Oren Zuckerman. “In-Car Game Design for Children: Child vs. Parent Perspective.” In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC ’13*, 112–19. New York, New York: ACM Press, 2013. https://doi.org/10.1145/2485760.2485768.

117. Horn, Michael S., Sarah AlSulaiman, and Jaime Koh. “Translating Roberto to Omar: Computational Literacy, Stickerbooks, and Cultural Forms.” In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC ’13*, 120–27. New York, New York: ACM Press, 2013. https://doi.org/10.1145/2485760.2485773.

118. Hourcade, Juan Pablo, Martha Driessnack, and Kelsey E. Huebner. “Supporting Face-to-Face Communication between Clinicians and Children with Chronic Headaches through a Zoomable Multi-Touch App.” In *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems - CHI ’12*, 2609. Austin, Texas, USA: ACM Press, 2012. https://doi.org/10.1145/2207676.2208651.

119. Hourcade, Juan Pablo, Keith B. Perry, and Aditya Sharma. “PointAssist: Helping Four Year Olds Point with Ease.” In *Proceedings of the 7th International Conference on Interaction Design and Children - IDC ’08*, 202. Chicago, Illinois: ACM Press, 2008. https://doi.org/10.1145/1463689.1463757.

120. Howland, Kate, Judith Good, and Benedict du Boulay. “Narrative Support for Young Game Designers’ Writing.” In *Proceedings of the 14th International Conference on Interaction Design and Children - IDC ’15*, 178–87. Boston, Massachusetts: ACM Press, 2015. https://doi.org/10.1145/2771839.2771858.

121. Huang, Kevin, Jesse Smith, Kimberly Spreen, and Mary Frances Jones. “Breaking the Sound Barrier: Designing an Interactive Tool for Language Acquisition in Preschool Deaf Children.” In *Proceedings of the 7th International Conference on Interaction Design and Children - IDC ’08*, 210. Chicago, Illinois: ACM Press, 2008. https://doi.org/10.1145/1463689.1463758.

122. Hudson, Nathaniel, Benjamin Lafreniere, Parmit K. Chilana, and Tovi Grossman. “Investigating How Online Help and Learning Resources Support Children’s Use of 3D Design Software.” In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI ’18*, 1–14. Montreal QC, Canada: ACM Press, 2018. https://doi.org/10.1145/3173574.3173831.

123. Ibrahim, Seray B., Asimina Vasalou, and Michael Clarke. “Design Opportunities for AAC and Children with Severe Speech and Physical Impairments.” In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI ’18*, 1–13. Montreal QC, Canada: ACM Press, 2018. https://doi.org/10.1145/3173574.3173801.

124. Iivari, Netta, Marianne Kinnula, Leena Kuure, and Tonja Molin-Juustila. “Video Diary as a Means for Data Gathering with Children – Encountering Identities in the Making.” *International Journal of Human-Computer Studies* 72, no. 5 (May 2014): 507–21. https://doi.org/10.1016/j.ijhcs.2014.02.003.

125. Iivari, Netta, Marianne Kinnula, and Tonja Molin-Juustila. “You Have to Start Somewhere: Initial Meanings Making in a Design and Making Project.” In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 80–92. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202742.

126. Ioannidi, Danai, Emmanouil Zidianakis, Margherita Antona, and Constantine Stephanidis. “Designing Games for Children with Developmental Disabilities in Ambient Intelligence Environments.” *International Journal of Child-Computer Interaction* 11 (January 2017): 40–49. https://doi.org/10.1016/j.ijcci.2016.10.008.

127. Iversen, Ole Sejer, Rachel Charlotte Smith, and Christian Dindler. “Child as Protagonist: Expanding the Role of Children in Participatory Design.” In *Proceedings of the 2017*

Conference on Interaction Design and Children, 27–37. Stanford California USA: ACM, 2017. https://doi.org/10.1145/3078072.3079725.

128. Jamil, Izdihar, Calkin Suero Montero, Mark Perry, Kenton O'Hara, Abhijit Karnik, Kaisa Pihlainen, Mark T. Marshall, Swathi Jha, Sanjay Gupta, and Sriram Subramanian. "Collaborating around Digital Tabletops: Children's Physical Strategies from India, the UK and Finland." *ACM Transactions on Computer-Human Interaction* 24, no. 3 (July 22, 2017): 1–30. https://doi.org/10.1145/3058551.
129. Jensen, Camilla Nørgaard, Winslow Burleson, and John Sadauskas. "Fostering Early Literacy Skills in Children's Libraries: Opportunities for Embodied Cognition and Tangible Technologies." In *Proceedings of the 11th International Conference on Interaction Design and Children - IDC '12*, 50. Bremen, Germany: ACM Press, 2012. https://doi.org/10.1145/2307096.2307103.
130. Johnson, Kristina T., and Rosalind W. Picard. "SPRING: Customizable, Motivation-Driven Technology for Children with Autism or Neurodevelopmental Differences." In *Proceedings of the 2017 Conference on Interaction Design and Children*, 149–58. Stanford California USA: ACM, 2017. https://doi.org/10.1145/3078072.3079718.
131. Jones, Susan, Lynne Hall, and Janette Hilton. "Exploring Children's 'indexical Encounter' with Real and Digitised Archive Photographs Using Tablet and Large Flat Screen Technologies." In *Proceedings of the 11th International Conference on Interaction Design and Children - IDC '12*, 30. Bremen, Germany: ACM Press, 2012. https://doi.org/10.1145/2307096.2307101.
132. Jurdi, Sandra, Fernando Garcia-Sanjuan, Vicente Nacher, and Javier Jaen. "Children's Acceptance of a Collaborative Problem Solving Game Based on Physical Versus Digital Learning Spaces." *Interacting with Computers* 30, no. 3 (May 1, 2018): 187–206. https://doi.org/10.1093/iwc/iwy006.
133. Kafai, Yasmin, and Veena Vasudevan. "Hi-Lo Tech Games: Crafting, Coding and Collaboration of Augmented Board Games by High School Youth." In *Proceedings of the 14th International Conference on Interaction Design and Children - IDC '15*, 130–39. Boston, Massachusetts: ACM Press, 2015. https://doi.org/10.1145/2771839.2771853.
134. Kammer, Dietrich, René Dang, Juliane Steinhauf, and Rainer Groh. "Investigating Interaction with Tabletops in Kindergarten Environments." In *Proceedings of the 2014 Conference on Interaction Design and Children - IDC '14*, 57–66. Aarhus, Denmark: ACM Press, 2014. https://doi.org/10.1145/2593968.2593975.

135. Kane, Shaun K., Varsha Koushik, and Annika Muehlbradt. “Bonk: Accessible Programming for Accessible Audio Games.” In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 132–42. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202754.

136. Kang, Seokbin, Leyla Norooz, Elizabeth Bonsignore, Virginia Byrne, Tamara Clegg, and Jon E. Froehlich. “PrototypAR: Prototyping and Simulating Complex Systems with Paper Craft and Augmented Reality.” In *Proceedings of the Interaction Design and Children on ZZZ - IDC ’19*, 253–66. Boise, ID, USA: ACM Press, 2019. https://doi.org/10.1145/3311927.3323135.

137. Katterfeldt, Eva-Sophie, Nadine Dittert, and Heidi Schelhowe. “EduWear: Smart Textiles as Ways of Relating Computing Technology to Everyday Life.” In *Proceedings of the 8th International Conference on Interaction Design and Children - IDC ’09*, 9. Como, Italy: ACM Press, 2009. https://doi.org/10.1145/1551788.1551791.

138. Kazemitabaar, Majeed, Jason McPeak, Alexander Jiao, Liang He, Thomas Outing, and Jon E. Froehlich. “MakerWear: A Tangible Approach to Interactive Wearable Creation for Children.” In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 133–45. Denver Colorado USA: ACM, 2017. https://doi.org/10.1145/3025453.3025887.

139. Keifert, Danielle, Christine Lee, Maggie Dahn, Randy Illum, David DeLiema, Noel Enyedy, and Joshua Danish. “Agency, Embodiment, & Affect During Play in a Mixed-Reality Learning Environment.” In *Proceedings of the 2017 Conference on Interaction Design and Children*, 268–77. Stanford California USA: ACM, 2017. https://doi.org/10.1145/3078072.3079731.

140. Kelly, Annie, Lila Finch, Monica Bolles, and R. Benjamin Shapiro. “BlockyTalky: New Programmable Tools to Enable Students’ Learning Networks.” *International Journal of Child-Computer Interaction* 18 (November 2018): 8–18. https://doi.org/10.1016/j.ijcci.2018.03.004.

141. Khowaja, Kamran, and Siti Salwah Salim. “Serious Game for Children with Autism to Learn Vocabulary: An Experimental Evaluation.” *International Journal of Human–Computer Interaction* 35, no. 1 (January 2, 2019): 1–26. https://doi.org/10.1080/10447318.2017.1420006.

142. Kim, Katherine K., Holly C. Logan, Edmund Young, and Christina M. Sabee. “Youth-Centered Design and Usage Results of the iN Touch Mobile Self-Management Program for Overweight/Obesity.” *Personal and Ubiquitous Computing* 19, no. 1 (January 2015): 59–68. https://doi.org/10.1007/s00779-014-0808-x.

143. Kory-Westlund, Jacqueline M., and Cynthia Breazeal. “Assessing Children’s Perceptions and Acceptance of a Social Robot.” In *Proceedings of the Interaction Design and Children on ZZZ - IDC ’19*, 38–50. Boise, ID, USA: ACM Press, 2019. https://doi.org/10.1145/3311927.3323143.

144. Kumar, Priya, Jessica Vitak, Marshini Chetty, Tamara L. Clegg, Jonathan Yang, Brenna McNally, and Elizabeth Bonsignore. “Co-Designing Online Privacy-Related Games and Stories with Children.” In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 67–79. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202735.

145. Kyfonidis, Charalampos, and Marilyn Lennon. “Making Diabetes Education Interactive: Tangible Educational Toys for Children with Type-1 Diabetes.” In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI ’19*, 1–12. Glasgow, Scotland Uk: ACM Press, 2019. https://doi.org/10.1145/3290605.3300671.

146. Kynigos, Chronis, and Nikoleta Yiannoutsou. “Children Challenging the Design of Half-Baked Games: Expressing Values through the Process of Game Modding.” *International Journal of Child-Computer Interaction* 17 (September 2018): 16–27. https://doi.org/10.1016/j.ijcci.2018.04.001.

147. Lamberty, K. K., Stephen Adams, Jason Biatek, Katherine Froiland, and Jay Lapham. “Using a Large Display in the Periphery to Support Children Learning through Design.” In *Proceedings of the 10th International Conference on Interaction Design and Children - IDC ’11*, 62–71. Ann Arbor, Michigan: ACM Press, 2011. https://doi.org/10.1145/1999030.1999038.

148. Landry, Pascal, Joseph Minsky, Marta Castañer, Oleguer Camerino, Rosa Rodriguez-Arregui, Enric Ormo, and Narcis Pares. “Design Strategy to Stimulate a Diversity of Motor Skills for an Exergame Addressed to Children.” In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC ’13*, 84–91. New York, New York: ACM Press, 2013. https://doi.org/10.1145/2485760.2485781.

149. Leduc-Mills, Ben, and Michael Eisenberg. “The UCube: A Child-Friendly Device for Introductory Three-Dimensional Design.” In *Proceedings of the 10th International Conference on Interaction Design and Children - IDC ’11*, 72–80. Ann Arbor, Michigan: ACM Press, 2011. https://doi.org/10.1145/1999030.1999039.

150. Lee, Victor R., Joel R. Drake, Ryan Cain, and Jeffrey Thayne. “Opportunistic Uses of the Traditional School Day through Student Examination of Fitbit Activity Tracker Data.” In *Proceedings of the 14th International Conference on Interaction Design and Children - IDC*

’15, 209–18. Boston, Massachusetts: ACM Press, 2015. https://doi.org/10.1145/2771839.2771861.

151. Leite, Iolanda, André Pereira, and Jill Fain Lehman. “Persistent Memory in Repeated Child-Robot Conversations.” In *Proceedings of the 2017 Conference on Interaction Design and Children*, 238–47. Stanford California USA: ACM, 2017. https://doi.org/10.1145/3078072.3079728.

152. Leong, Zeina Atrash, and Michael S. Horn. “Waiting for Learning: Designing Interactive Education Materials for Patient Waiting Areas.” In *Proceedings of the 2014 Conference on Interaction Design and Children - IDC ’14*, 145–53. Aarhus, Denmark: ACM Press, 2014. https://doi.org/10.1145/2593968.2593970.

153. Lim, Bohyeon, Yvonne Rogers, and Neil Sebire. “Designing to Distract: Can Interactive Technologies Reduce Visitor Anxiety in a Children’s Hospital Setting?” *ACM Transactions on Computer-Human Interaction* 26, no. 2 (April 28, 2019): 1–19. https://doi.org/10.1145/3301427.

154. Lindberg, Susanne, Pontus Wärnestål, Jens Nygren, and Petra Svedberg. “Designing Digital Peer Support for Children: Design Patterns for Social Interaction.” In *Proceedings of the 2014 Conference on Interaction Design and Children - IDC ’14*, 47–56. Aarhus, Denmark: ACM Press, 2014. https://doi.org/10.1145/2593968.2593972.

155. Liszio, Stefan, and Maic Masuch. “Virtual Reality MRI: Playful Reduction of Children’s Anxiety in MRI Exams.” In *Proceedings of the 2017 Conference on Interaction Design and Children*, 127–36. Stanford California USA: ACM, 2017. https://doi.org/10.1145/3078072.3079713.

156. Lovato, Silvia B., Anne Marie Piper, and Ellen A. Wartella. “Hey Google, Do Unicorns Exist?: Conversational Agents as a Path to Answers to Children’s Questions.” In *Proceedings of the Interaction Design and Children on ZZZ - IDC ’19*, 301–13. Boise, ID, USA: ACM Press, 2019. https://doi.org/10.1145/3311927.3323150.

157. Lyons, Leilah, Michael Tissenbaum, Matthew Berland, Rebecca Eydt, Lauren Wielgus, and Adam Mechtley. “Designing Visible Engineering: Supporting Tinkering Performances in Museums.” In *Proceedings of the 14th International Conference on Interaction Design and Children - IDC ’15*, 49–58. Boston, Massachusetts: ACM Press, 2015. https://doi.org/10.1145/2771839.2771845.

158. Ma, Yudan, Tilde Bekker, Xipei Ren, Jun Hu, and Steven Vos. “Effects of Playful Audio Augmentation on Teenagers’ Motivations in Cooperative Physical Play.” In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 43–54. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202729.

159. Maldonado, Heidy, and Ariel Zekelman. “Designing Tangible ABCs: Fröbel’s Sticks and Rings for the 21st Century.” In *Proceedings of the Interaction Design and Children on ZZZ - IDC ’19*, 326–33. Boise, ID, USA: ACM Press, 2019. https://doi.org/10.1145/3311927.3323123.

160. Malinverni, Laura, Joan MoraGuiard, Vanesa Padillo, MariaAngeles Mairena, Amaia Hervás, and Narcis Pares. “Participatory Design Strategies to Enhance the Creative Contribution of Children with Special Needs.” In *Proceedings of the 2014 Conference on Interaction Design and Children - IDC ’14*, 85–94. Aarhus, Denmark: ACM Press, 2014. https://doi.org/10.1145/2593968.2593981.

161. Malinverni, Laura, Brenda López Silva, and Narcís Parés. “Impact of Embodied Interaction on Learning Processes: Design and Analysis of an Educational Application Based on Physical Activity.” In *Proceedings of the 11th International Conference on Interaction Design and Children - IDC ’12*, 60. Bremen, Germany: ACM Press, 2012. https://doi.org/10.1145/2307096.2307104.

162. Malinverni, Laura, Cristina Valero, Marie-Monique Schaper, and Narcis Pares. “A Conceptual Framework to Compare Two Paradigms of Augmented and Mixed Reality Experiences.” In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 7–18. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202750.

163. Mandryk, Regan L., Shane Dielschneider, Michael R. Kalyn, Christopher P. Bertram, Michael Gaetz, Andre Doucette, Brett A. Taylor, Alison Pritchard Orr, and Kathy Keiver. “Games as Neurofeedback Training for Children with FASD.” In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC ’13*, 165–72. New York, New York: ACM Press, 2013. https://doi.org/10.1145/2485760.2485762.

164. Mansor, Evi Indriasari, Antonella De Angeli, and Oscar de Bruijn. “The Fantasy Table.” In *Proceedings of the 8th International Conference on Interaction Design and Children - IDC ’09*, 70. Como, Italy: ACM Press, 2009. https://doi.org/10.1145/1551788.1551801.

165. Maqsood, Sana, Christine Mekhail, and Sonia Chiasson. “A Day in the Life of Jos: A Web-Based Game to Increase Children’s Digital Literacy.” In *Proceedings of the 17th ACM*

Conference on Interaction Design and Children, 241–52. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202753.

166. Marco, Javier, Eva Cerezo, and Sandra Baldassarri. “Bringing Tabletop Technology to All: Evaluating a Tangible Farm Game with Kindergarten and Special Needs Children.” *Personal and Ubiquitous Computing* 17, no. 8 (December 2013): 1577–91. https://doi.org/10.1007/s00779-012-0522-5.

167. Martin-Niedecken, Anna Lisa. “Designing for Bodily Interplay: Engaging with the Adaptive Social Exertion Game ‘Plunder Planet.’” In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 19–30. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202740.

168. Matthews, Mark, and Gavin Doherty. “In the Mood: Engaging Teenagers in Psychotherapy Using Mobile Phones.” In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI ’11*, 2947. Vancouver, BC, Canada: ACM Press, 2011. https://doi.org/10.1145/1978942.1979379.

169. Matviienko, Andrii, Swamy Ananthanarayan, Abdallah El Ali, Wilko Heuten, and Susanne Boll. “NaviBike: Comparing Unimodal Navigation Cues for Child Cyclists.” In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI ’19*, 1–12. Glasgow, Scotland Uk: ACM Press, 2019. https://doi.org/10.1145/3290605.3300850.

170. Mazzone, Emanuela, Netta Iivari, Ruut Tikkanen, Janet C. Read, and Russell Beale. “Considering Context, Content, Management, and Engagement in Design Activities with Children.” In *Proceedings of the 9th International Conference on Interaction Design and Children - IDC ’10*, 108. Barcelona, Spain: ACM Press, 2010. https://doi.org/10.1145/1810543.1810556.

171. McBeath, Jasmine K., Richard P. Durán, and Danielle B. Harlow. “Not My Gumdrop Buttons!: Youth Tool Use in Designing an Electronic Shrek-Themed Bean Bag Toss.” In *Proceedings of the 2017 Conference on Interaction Design and Children*, 61–72. Stanford California USA: ACM, 2017. https://doi.org/10.1145/3078072.3079721.

172. McNally, Brenna, Mona Leigh Guha, Leyla Norooz, Emily Rhodes, and Leah Findlater. “Incorporating Peephole Interactions into Children’s Second Language Learning Activities on Mobile Devices.” In *Proceedings of the 2014 Conference on Interaction Design and Children - IDC ’14*, 115–24. Aarhus, Denmark: ACM Press, 2014. https://doi.org/10.1145/2593968.2593982.

173. McReynolds, Emily, Sarah Hubbard, Timothy Lau, Aditya Saraf, Maya Cakmak, and Franziska Roesner. “Toys That Listen: A Study of Parents, Children, and Internet-Connected Toys.” In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 5197–5207. Denver Colorado USA: ACM, 2017. https://doi.org/10.1145/3025453.3025735.

174. McRoberts, Sarah, Ye Yuan, Kathleen Watson, and Svetlana Yarosh. “Behind the Scenes: Design, Collaboration, and Video Creation with Youth.” In *Proceedings of the Interaction Design and Children on ZZZ - IDC ’19*, 173–84. Boise, ID, USA: ACM Press, 2019. https://doi.org/10.1145/3311927.3323134.

175. Metatla, Oussama, Emanuela Maggioni, Clare Cullen, and Marianna Obrist. “‘Like Popcorn’: Crossmodal Correspondences Between Scents, 3D Shapes and Emotions in Children.” In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI ’19*, 1–13. Glasgow, Scotland Uk: ACM Press, 2019. https://doi.org/10.1145/3290605.3300689.

176. Meyns, Pieter, Judith van der Spank, Hanne Capiau, Lieve De Cock, Eline Van Steirteghem, Ruth Van der Looven, and Hilde Van Waelvelde. “Do a Humanoid Robot and Music Increase the Motivation to Perform Physical Activity? A Quasi-Experimental Cohort in Typical Developing Children and Preliminary Findings in Hospitalized Children in Neutropenia.” *International Journal of Human-Computer Studies* 122 (February 2019): 90–102. https://doi.org/10.1016/j.ijhcs.2018.07.010.

177. Michaelis, Joseph E., and Bilge Mutlu. “Supporting Interest in Science Learning with a Social Robot.” In *Proceedings of the Interaction Design and Children on ZZZ - IDC ’19*, 71–82. Boise, ID, USA: ACM Press, 2019. https://doi.org/10.1145/3311927.3323154.

178. Mills, Kelly, Elizabeth Bonsignore, Tamara Clegg, June Ahn, Jason Yip, Daniel Pauw, Lautaro Cabrera, Kenna Hernly, and Caroline Pitt. “Designing to Illuminate Children’s Scientific Funds of Knowledge through Social Media Sharing.” In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 266–77. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202737.

179. Moher, Tom. “Putting Interference to Work in the Design of a Whole-Class Learning Activity.” In *Proceedings of the 8th International Conference on Interaction Design and Children - IDC ’09*, 115. Como, Italy: ACM Press, 2009. https://doi.org/10.1145/1551788.1551808.

180. Mora-Guiard, Joan, Ciera Crowell, Narcis Pares, and Pamela Heaton. “Sparking Social Initiation Behaviors in Children with Autism through Full-Body Interaction.” *International*

Journal of Child-Computer Interaction 11 (January 2017): 62–71. https://doi.org/10.1016/j.ijcci.2016.10.006.

181. MoraGuiard, Joan, and Narcis Pares. "'Child as the Measure of All Things': The Body as a Referent in Designing a Museum Exhibit to Understand the Nanoscale." In *Proceedings of the 2014 Conference on Interaction Design and Children - IDC '14*, 27–36. Aarhus, Denmark: ACM Press, 2014. https://doi.org/10.1145/2593968.2593985.
182. Moraveji, Neema, Kori Inkpen, Ed Cutrell, and Ravin Balakrishnan. "A Mischief of Mice: Examining Children's Performance in Single Display Groupware Systems with 1 to 32 Mice." In *Proceedings of the 27th International Conference on Human Factors in Computing Systems - CHI 09*, 2157. Boston, MA, USA: ACM Press, 2009. https://doi.org/10.1145/1518701.1519030.
183. Nacher, Vicente, Doris Cáliz, Javier Jaen, and Loïc Martínez. "Examining the Usability of Touch Screen Gestures for Children With Down Syndrome." *Interacting with Computers* 30, no. 3 (May 1, 2018): 258–72. https://doi.org/10.1093/iwc/iwy011.
184. Nacher, Vicente, Fernando Garcia-Sanjuan, and Javier Jaen. "Evaluating Simultaneous Visual Instructions with Kindergarten Children on Touchscreen Devices." *International Journal of Human–Computer Interaction* 36, no. 1 (January 2, 2020): 41–54. https://doi.org/10.1080/10447318.2019.1597576.
185. Ngai, Grace, Stephen C.F. Chan, Vincent T.Y. Ng, Joey C.Y. Cheung, Sam S.S. Choy, Winnie W.Y. Lau, and Jason T.P. Tse. "I*CATch: A Scalable Plug-n-Play Wearable Computing Framework for Novices and Children." In *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10*, 443. Atlanta, Georgia, USA: ACM Press, 2010. https://doi.org/10.1145/1753326.1753393.
186. Nielsen, Rune, Jonas Fritsch, Kim Halskov, and Martin Brynskov. "Out of the Box: Exploring the Richness of Children's Use of an Interactive Table." In *Proceedings of the 8th International Conference on Interaction Design and Children - IDC '09*, 61. Como, Italy: ACM Press, 2009. https://doi.org/10.1145/1551788.1551800.
187. Nouwen, Marije, Selina Schepers, Karen Mouws, Karin Slegers, Niek Kosten, and Pieter Duysburgh. "Designing an Educational Music Game: What If Children Were Calling the Tune?" *International Journal of Child-Computer Interaction* 9–10 (December 2016): 20–32. https://doi.org/10.1016/j.ijcci.2016.10.001.

188. Novellis, Francesco, and Tom Moher. “How Real Is ‘Real Enough’?: Designing Artifacts and Procedures for Embodied Simulations of Science Practices.” In *Proceedings of the 10th International Conference on Interaction Design and Children - IDC ’11*, 90–98. Ann Arbor, Michigan: ACM Press, 2011. https://doi.org/10.1145/1999030.1999041.

189. Odom, William, John Zimmerman, and Jodi Forlizzi. “Teenagers and Their Virtual Possessions: Design Opportunities and Issues.” In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI ’11*, 1491. Vancouver, BC, Canada: ACM Press, 2011. https://doi.org/10.1145/1978942.1979161.

190. Ofer, Netta, Idan David, Hadas Erel, and Oren Zuckerman. “Coding for Outdoor Play: A Coding Platform for Children to Invent and Enhance Outdoor Play Experiences.” In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI ’19*, 1–12. Glasgow, Scotland Uk: ACM Press, 2019. https://doi.org/10.1145/3290605.3300394.

191. Ouchi, Hisakazu, Yoshifumi Nishida, Ilwoong Kim, Yoichi Motomura, and Hiroshi Mizoguchi. “Detecting and Modeling Play Behavior Using Sensor-Embedded Rock-Climbing Equipment.” In *Proceedings of the 9th International Conference on Interaction Design and Children - IDC ’10*, 118. Barcelona, Spain: ACM Press, 2010. https://doi.org/10.1145/1810543.1810557.

192. Pantoja, Luiza Superti, Kyle Diederich, Liam Crawford, and Juan Pablo Hourcade. “Voice Agents Supporting High-Quality Social Play.” In *Proceedings of the Interaction Design and Children on ZZZ - IDC ’19*, 314–25. Boise, ID, USA: ACM Press, 2019. https://doi.org/10.1145/3311927.3323151.

193. Papavlasopoulou, Sofia, Kshitij Sharma, Michail Giannakos, and Letizia Jaccheri. “Using Eye-Tracking to Unveil Differences Between Kids and Teens in Coding Activities.” In *Proceedings of the 2017 Conference on Interaction Design and Children*, 171–81. Stanford California USA: ACM, 2017. https://doi.org/10.1145/3078072.3079740.

194. Park, Tina, Peter Chira, Kimberly Miller, and Lisa Nugent. “Living Profiles: An Example of User-Centered Design in Developing a Teen-Oriented Personal Health Record.” *Personal and Ubiquitous Computing* 19, no. 1 (January 2015): 69–77. https://doi.org/10.1007/s00779-014-0812-1.

195. Qiu, Kanjun, Leah Buechley, Edward Baafi, and Wendy Dubow. “A Curriculum for Teaching Computer Science through Computational Textiles.” In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC ’13*, 20–27. New York, New York: ACM Press, 2013. https://doi.org/10.1145/2485760.2485787.

196. Raffle, Hayes, Mirjana Spasojevic, Rafael Ballagas, Glenda Revelle, Hiroshi Horii, Sean Follmer, Janet Go, Emily Reardon, Koichi Mori, and Joseph Kaye. "Family Story Play: Reading with Young Children (and Elmo) over a Distance." In *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10*, 1583. Atlanta, Georgia, USA: ACM Press, 2010. https://doi.org/10.1145/1753326.1753563.

197. Raya, Rafael, Ramón Ceres, Javier O. Roa, and Eduardo Rocon. "Assessment of the Involuntary Motion of Children with Motor Impairments to Improve the Accessibility of an Inertial Interface." In *Proceedings of the 9th International Conference on Interaction Design and Children - IDC '10*, 128. Barcelona, Spain: ACM Press, 2010. https://doi.org/10.1145/1810543.1810558.

198. Read, J. C., D. Fitton, and M. Horton. "Theatre, PlayDoh and Comic Strips: Designing Organic User Interfaces with Young Adolescent and Teenage Participants." *Interacting with Computers* 25, no. 2 (March 1, 2013): 183–98. https://doi.org/10.1093/iwc/iws016.

199. Read, Janet C., Daniel Fitton, and Matthew Hortton. "Giving Ideas an Equal Chance: Inclusion and Representation in Participatory Design with Children." In *Proceedings of the 2014 Conference on Interaction Design and Children - IDC '14*, 105–14. Aarhus, Denmark: ACM Press, 2014. https://doi.org/10.1145/2593968.2593986.

200. Rice, Mark, Wah Pheow Tan, Jeremy Ong, Lih Jie Yau, Marcus Wan, and Jamie Ng. "The Dynamics of Younger and Older Adult's Paired Behavior When Playing an Interactive Silhouette Game." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, 1081. Paris, France: ACM Press, 2013. https://doi.org/10.1145/2470654.2466138.

201. Rick, Jochen, Phyllis Francois, Bob Fields, Rowanne Fleck, Nicola Yuill, and Amanda Carr. "Lo-Fi Prototyping to Design Interactive-Tabletop Applications for Children." In *Proceedings of the 9th International Conference on Interaction Design and Children - IDC '10*, 138. Barcelona, Spain: ACM Press, 2010. https://doi.org/10.1145/1810543.1810559.

202. Rick, Jochen, Amanda Harris, Paul Marshall, Rowanne Fleck, Nicola Yuill, and Yvonne Rogers. "Children Designing Together on a Multi-Touch Tabletop: An Analysis of Spatial Orientation and User Interactions." In *Proceedings of the 8th International Conference on Interaction Design and Children - IDC '09*, 106. Como, Italy: ACM Press, 2009. https://doi.org/10.1145/1551788.1551807.

203. Rick, Jochen, Paul Marshall, and Nicola Yuill. “Beyond One-Size-Fits-All: How Interactive Tabletops Support Collaborative Learning.” In *Proceedings of the 10th International Conference on Interaction Design and Children - IDC ’11*, 109–17. Ann Arbor, Michigan: ACM Press, 2011. https://doi.org/10.1145/1999030.1999043.

204. Roberts, Jessica, Leilah Lyons, Francesco Cafaro, and Rebecca Eydt. “Interpreting Data from within: Supporting Humandata Interaction in Museum Exhibits through Perspective Taking.” In *Proceedings of the 2014 Conference on Interaction Design and Children - IDC ’14*, 7–16. Aarhus, Denmark: ACM Press, 2014. https://doi.org/10.1145/2593968.2593974.

205. Robertson, Judy, Andrew Macvean, and Katy Howland. “Embedding Technology in the Classroom: The Train the Teacher Model.” In *Proceedings of the 11th International Conference on Interaction Design and Children - IDC ’12*, 20. Bremen, Germany: ACM Press, 2012. https://doi.org/10.1145/2307096.2307100.

206. Roper, Tessa, Laura Millen Dutka, Sue Cobb, and Harshada Patel. “Collaborative Virtual Environment to Facilitate Game Design Evaluation with Children with ASC.” *International Journal of Human–Computer Interaction* 35, no. 8 (May 9, 2019): 692–705. https://doi.org/10.1080/10447318.2018.1550179.

207. Rubegni, Elisa, and Monica Landoni. “Fiabot!: Design and Evaluation of a Mobile Storytelling Application for Schools.” In *Proceedings of the 2014 Conference on Interaction Design and Children - IDC ’14*, 165–74. Aarhus, Denmark: ACM Press, 2014. https://doi.org/10.1145/2593968.2593979.

208. Ryokai, Kimiko, Hayes Raffle, and Robert Kowalski. “StoryFaces: Pretend-Play with Ebooks to Support Social-Emotional Storytelling.” In *Proceedings of the 11th International Conference on Interaction Design and Children - IDC ’12*, 125. Bremen, Germany: ACM Press, 2012. https://doi.org/10.1145/2307096.2307111.

209. Ryokai, Kimiko, Peiqi Su, Eungchan Kim, and Bob Rollins. “EnergyBugs: Energy Harvesting Wearables for Children.” In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI ’14*, 1039–48. Toronto, Ontario, Canada: ACM Press, 2014. https://doi.org/10.1145/2556288.2557225.

210. Sadka, Ofir, Hadas Erel, Andrey Grishko, and Oren Zuckerman. “Tangible Interaction in Parent-Child Collaboration: Encouraging Awareness and Reflection.” In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 157–69. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202746.

211. Sapounidis, Theodosios, Stavros Demetriadis, Pantelis M. Papadopoulos, and Dimitrios Stamovlasis. “Tangible and Graphical Programming with Experienced Children: A Mixed Methods Analysis.” *International Journal of Child-Computer Interaction* 19 (March 2019): 67–78. https://doi.org/10.1016/j.ijcci.2018.12.001.

212. Saturno, Cleiton Eduardo, Alejandro Rafael Garcia Ramirez, Mauro José Conte, Mísia Farhat, and Elaine Carmelita Piucco. “An Augmentative and Alternative Communication Tool for Children and Adolescents with Cerebral Palsy.” *Behaviour & Information Technology* 34, no. 6 (June 3, 2015): 632–45. https://doi.org/10.1080/0144929X.2015.1019567.

213. Schaper, Marie-Monique, Maria Santos, and Narcis Pares. “Orchestrating Experts’ Assumptions and Children’s Values in the Design of Virtual Heritage Experiences.” *International Journal of Child-Computer Interaction* 17 (September 2018): 5–15. https://doi.org/10.1016/j.ijcci.2018.02.001.

214. Scharf, Florian, Thomas Winkler, Claudia Hahn, Christian Wolters, and Michael Herczeg. “Tangicons 3.0: An Educational Non-Competitive Collaborative Game.” In *Proceedings of the 11th International Conference on Interaction Design and Children - IDC ’12*, 144. Bremen, Germany: ACM Press, 2012. https://doi.org/10.1145/2307096.2307113.

215. Scharf, Florian, Thomas Winkler, and Michael Herczeg. “Tangicons: Algorithmic Reasoning in a Collaborative Game for Children in Kindergarten and First Class.” In *Proceedings of the 7th International Conference on Interaction Design and Children - IDC ’08*, 242. Chicago, Illinois: ACM Press, 2008. https://doi.org/10.1145/1463689.1463762.

216. Segura, Elena Márquez, Henriette Cramer, Paulo Fontaínha Gomes, Stina Nylander, and Ana Paiva. “Revive!: Reactions to Migration between Different Embodiments When Playing with Robotic Pets.” In *Proceedings of the 11th International Conference on Interaction Design and Children - IDC ’12*, 88. Bremen, Germany: ACM Press, 2012. https://doi.org/10.1145/2307096.2307107.

217. Sengupta, Pratim, and Amy Voss Farris. “Learning Kinematics in Elementary Grades Using Agent-Based Computational Modeling: A Visual Programming-Based Approach.” In *Proceedings of the 11th International Conference on Interaction Design and Children - IDC ’12*, 78. Bremen, Germany: ACM Press, 2012. https://doi.org/10.1145/2307096.2307106.

218. Sharma, Kshitij, Sofia Papavlasopoulou, and Michail Giannakos. “Joint Emotional State of Children and Perceived Collaborative Experience in Coding Activities.” In *Proceedings of the*

Interaction Design and Children on ZZZ - IDC ’19, 133–45. Boise, ID, USA: ACM Press, 2019. https://doi.org/10.1145/3311927.3323145.

219. Sheriff, Aviv, Rona Sadan, Yasmin Keats, and Oren Zuckerman. “From Smart Homes to Smart Kids: Design Research for CataKit.” In *Proceedings of the 2017 Conference on Interaction Design and Children*, 159–69. Stanford California USA: ACM, 2017. https://doi.org/10.1145/3078072.3079729.

220. Shimoda, Todd, Barbara White, Marcela Borge, and John Frederiksen. “Designing for Science Learning and Collaborative Discourse.” In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC ’13*, 247–56. New York, New York: ACM Press, 2013. https://doi.org/10.1145/2485760.2485782.

221. Sim, Gavin, Brendan Cassidy, and Janet C. Read. “Understanding the Fidelity Effect When Evaluating Games with Children.” In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC ’13*, 193–200. New York, New York: ACM Press, 2013. https://doi.org/10.1145/2485760.2485769.

222. Sim, Gavin, and Matthew Horton. “Investigating Children’s Opinions of Games: Fun Toolkit vs. This or That.” In *Proceedings of the 11th International Conference on Interaction Design and Children - IDC ’12*, 70. Bremen, Germany: ACM Press, 2012. https://doi.org/10.1145/2307096.2307105.

223. Sim, Gavin, Marije Nouwen, Jorick Vissers, Matthew Horton, Karin Slegers, and Bieke Zaman. “Using the MemoLine to Capture Changes in User Experience over Time with Children.” *International Journal of Child-Computer Interaction* 8 (May 2016): 1–14. https://doi.org/10.1016/j.ijcci.2016.07.001.

224. Sipitakiat, Arnan, and Nusarin Nusen. “Robo-Blocks: Designing Debugging Abilities in a Tangible Programming System for Early Primary School Children.” In *Proceedings of the 11th International Conference on Interaction Design and Children - IDC ’12*, 98. Bremen, Germany: ACM Press, 2012. https://doi.org/10.1145/2307096.2307108.

225. Smith, Rachel C., Ole S. Iversen, Thomas Hjermitslev, and Aviaja B. Lynggaard. “Towards an Ecological Inquiry in Child-Computer Interaction.” In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC ’13*, 183–92. New York, New York: ACM Press, 2013. https://doi.org/10.1145/2485760.2485780.

226. Sobel, Kiley, Katie O’Leary, and Julie A. Kientz. “Maximizing Children’s Opportunities with Inclusive Play: Considerations for Interactive Technology Design.” In *Proceedings of the 14th*

International Conference on Interaction Design and Children - IDC ’15, 39–48. Boston, Massachusetts: ACM Press, 2015. https://doi.org/10.1145/2771839.2771844.

227. Sobel, Kiley, Kyle Rector, Susan Evans, and Julie A. Kientz. “Incloodle: Evaluating an Interactive Application for Young Children with Mixed Abilities.” In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 165–76. San Jose California USA: ACM, 2016. https://doi.org/10.1145/2858036.2858114.

228. Sonne, Tobias, Jörg Müller, Paul Marshall, Carsten Obel, and Kaj Grønbæk. “Changing Family Practices with Assistive Technology: MOBERO Improves Morning and Bedtime Routines for Children with ADHD.” In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 152–64. San Jose California USA: ACM, 2016. https://doi.org/10.1145/2858036.2858157.

229. Soute, Iris, Susanne Lagerström, and Panos Markopoulos. “Rapid Prototyping of Outdoor Games for Children in an Iterative Design Process.” In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC ’13*, 74–83. New York, New York: ACM Press, 2013. https://doi.org/10.1145/2485760.2485779.

230. Sysoev, Ivan, Anneli Hershman, Susan Fine, Claire Traweek, and Deb Roy. “SpeechBlocks: A Constructionist Early Literacy App.” In *Proceedings of the 2017 Conference on Interaction Design and Children*, 248–57. Stanford California USA: ACM, 2017. https://doi.org/10.1145/3078072.3079720.

231. Tarkan, Sureyya, Vibha Sazawal, Allison Druin, Evan Golub, Elizabeth M. Bonsignore, Greg Walsh, and Zeina Atrash. “Toque: Designing a Cooking-Based Programming Language for and with Children.” In *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI ’10*, 2417. Atlanta, Georgia, USA: ACM Press, 2010. https://doi.org/10.1145/1753326.1753692.

232. Toscos, Tammy, Kay Connelly, and Yvonne Rogers. “Best Intentions: Health Monitoring Technology and Children.” In *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems - CHI ’12*, 1431. Austin, Texas, USA: ACM Press, 2012. https://doi.org/10.1145/2207676.2208603.

233. Trory, Anthony, Kate Howland, and Judith Good. “Designing for Concreteness Fading in Primary Computing.” In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 278–88. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202748.

234. Tseng, Tiffany, Coram Bryant, and Paulo Blikstein. “Collaboration through Documentation: Automated Capturing of Tangible Constructions to Support Engineering Design.” In *Proceedings of the 10th International Conference on Interaction Design and Children - IDC ’11*, 118–26. Ann Arbor, Michigan: ACM Press, 2011. https://doi.org/10.1145/1999030.1999044.

235. Tsvyatkova, Damyanka, and Cristiano Storni. “Designing an Educational Interactive EBook for Newly Diagnosed Children with Type 1 Diabetes: Mapping a New Design Space.” *International Journal of Child-Computer Interaction* 19 (March 2019): 1–18. https://doi.org/10.1016/j.ijcci.2018.10.001.

236. Ucar, Erdem, Hakan Ustunel, Turhan Civelek, and Ilhan Umut. “Effects of Using a Force Feedback Haptic Augmented Simulation on the Attitudes of the Gifted Students towards Studying Chemical Bonds in Virtual Reality Environment.” *Behaviour & Information Technology* 36, no. 5 (May 4, 2017): 540–47. https://doi.org/10.1080/0144929X.2016.1264483.

237. Vacca, Ralph. “Bicultural: Examining Teenage Latinas’ Perspectives on Technologies for Emotional Support.” In *Proceedings of the 2017 Conference on Interaction Design and Children*, 117–26. Stanford California USA: ACM, 2017. https://doi.org/10.1145/3078072.3079742.

238. Vacca, Ralph. “Brokering Data: Co-Designing Technology with Latina Teens to Support Communication with Parents: Leveraging Cultural Practices of Latinx Youth through Co-Design.” In *Proceedings of the Interaction Design and Children on ZZZ - IDC ’19*, 197–207. Boise, ID, USA: ACM Press, 2019. https://doi.org/10.1145/3311927.3323142.

239. Van Mechelen, Maarten, Ann Laenen, Bieke Zaman, Bert Willems, and Vero Vanden Abeele. “Collaborative Design Thinking (CoDeT): A Co-Design Approach for High Child-to-Adult Ratios.” *International Journal of Human-Computer Studies* 130 (October 2019): 179–95. https://doi.org/10.1016/j.ijhcs.2019.06.013.

240. Van Mechelen, Maarten, Bieke Zaman, Ann Laenen, and Vero Vanden Abeele. “Challenging Group Dynamics in Participatory Design with Children: Lessons from Social Interdependence Theory.” In *Proceedings of the 14th International Conference on Interaction Design and Children - IDC ’15*, 219–28. Boston, Massachusetts: ACM Press, 2015. https://doi.org/10.1145/2771839.2771862.

241. Vatavu, Radu-Daniel, Gabriel Cramariuc, and Doina Maria Schipor. "Touch Interaction for Children Aged 3 to 6 Years: Experimental Findings and Relationship to Motor Skills." *International Journal of Human-Computer Studies* 74 (February 2015): 54–76. https://doi.org/10.1016/j.ijhcs.2014.10.007.

242. Virnes, Marjo, Erkki Sutinen, and Eija Kärnä-Lin. "How Children's Individual Needs Challenge the Design of Educational Robotics." In *Proceedings of the 7th International Conference on Interaction Design and Children - IDC '08*, 274. Chicago, Illinois: ACM Press, 2008. https://doi.org/10.1145/1463689.1463766.

243. Waddington, Jonathan, Conor Linehan, Kathrin Gerling, Kieran Hicks, and Timothy L. Hodgson. "Participatory Design of Therapeutic Video Games for Young People with Neurological Vision Impairment." In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 3533–42. Seoul, Republic of Korea: ACM Press, 2015. https://doi.org/10.1145/2702123.2702261.

244. Walsh, Greg, Robin Brewer, Asmi Joshi, Richelle Brown, Allison Druin, Mona Leigh Guha, Elizabeth Bonsignore, et al. "DisCo: A Co-Design Online Tool for Asynchronous Distributed Child and Adult Design Partners." In *Proceedings of the 11th International Conference on Interaction Design and Children - IDC '12*, 11. Bremen, Germany: ACM Press, 2012. https://doi.org/10.1145/2307096.2307099.

245. Walsh, Greg, and Elizabeth Foss. "A Case for Intergenerational Distributed Co-Design: The Online Kidsteam Example." In *Proceedings of the 14th International Conference on Interaction Design and Children - IDC '15*, 99–108. Boston, Massachusetts: ACM Press, 2015. https://doi.org/10.1145/2771839.2771850.

246. Wang, Danli, Cheng Zhang, and Hongan Wang. "T-Maze: A Tangible Programming Tool for Children." In *Proceedings of the 10th International Conference on Interaction Design and Children - IDC '11*, 127–35. Ann Arbor, Michigan: ACM Press, 2011. https://doi.org/10.1145/1999030.1999045.

247. Webster, Mary, Emma Foster, Rob Comber, Simon Bowen, Tim Cheetham, and Madeline Balaam. "Understanding the Lived Experience of Adolescents with Type 1 Diabetes: Opportunities for Design." In *Proceedings of the 14th International Conference on Interaction Design and Children - IDC '15*, 140–49. Boston, Massachusetts: ACM Press, 2015. https://doi.org/10.1145/2771839.2771854.

248. Weintrop, David, and Uri Wilensky. “Between a Block and a Typeface: Designing and Evaluating Hybrid Programming Environments.” In *Proceedings of the 2017 Conference on Interaction Design and Children*, 183–92. Stanford California USA: ACM, 2017. https://doi.org/10.1145/3078072.3079715.

249. Weintrop, David, and Uri Wilensky. “To Block or Not to Block, That Is the Question: Students’ Perceptions of Blocks-Based Programming.” In *Proceedings of the 14th International Conference on Interaction Design and Children - IDC ’15*, 199–208. Boston, Massachusetts: ACM Press, 2015. https://doi.org/10.1145/2771839.2771860.

250. Weisblatt, E. J., C. S. Langensiepen, B. Cook, C. Dias, K. Plaisted Grant, M. Dhariwal, M. S. Fairclough, et al. “A Tablet Computer-Assisted Motor and Language Skills Training Program to Promote Communication Development in Children with Autism: Development and Pilot Study.” *International Journal of Human–Computer Interaction* 35, no. 8 (May 9, 2019): 643–65. https://doi.org/10.1080/10447318.2018.1550176.

251. Westlund, Jacqueline M. Kory, Hae Won Park, Randi Williams, and Cynthia Breazeal. “Measuring Young Children’s Long-Term Relationships with Social Robots.” In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 207–18. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202732.

252. Wilson, Cara, Margot Brereton, Bernd Ploderer, and Laurianne Sitbon. “Co-Design Beyond Words: ‘Moments of Interaction’ with Minimally-Verbal Children on the Autism Spectrum.” In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI ’19*, 1–15. Glasgow, Scotland Uk: ACM Press, 2019. https://doi.org/10.1145/3290605.3300251.

253. Wilson, Cara, Margot Brereton, Bernd Ploderer, and Laurianne Sitbon. “MyWord: Enhancing Engagement, Interaction and Self-Expression with Minimally-Verbal Children on the Autism Spectrum through a Personal Audio-Visual Dictionary.” In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 106–18. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202755.

254. Wood, Matthew, Gavin Wood, and Madeline Balaam. “Sex Talk: Designing for Sexual Health with Adolescents.” In *Proceedings of the 2017 Conference on Interaction Design and Children*, 137–47. Stanford California USA: ACM, 2017. https://doi.org/10.1145/3078072.3079747.

255. Woodward, Julia, Zari McFadden, Nicole Shiver, Amir Ben-hayon, Jason C. Yip, and Lisa Anthony. “Using Co-Design to Examine How Children Conceptualize Intelligent Interfaces.” In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI ’18*, 1–14. Montreal QC, Canada: ACM Press, 2018. https://doi.org/10.1145/3173574.3174149.

256. Woodward, Julia, Alex Shaw, Annie Luc, Brittany Craig, Juthika Das, Phillip Hall, Akshay Holla, et al. “Characterizing How Interface Complexity Affects Children’s Touchscreen Interactions.” In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 1921–33. San Jose California USA: ACM, 2016. https://doi.org/10.1145/2858036.2858200.

257. Wyeth, Peta, and Ian MacColl. “Noising around: Investigations in Mobile Learning.” In *Proceedings of the 9th International Conference on Interaction Design and Children - IDC ’10*, 147. Barcelona, Spain: ACM Press, 2010. https://doi.org/10.1145/1810543.1810560.

258. Xie, Bo, Allison Druin, Jerry Fails, Sheri Massey, Evan Golub, Sonia Franckel, and Kiki Schneider. “Connecting Generations: Developing Co-Design Methods for Older Adults and Children.” *Behaviour & Information Technology* 31, no. 4 (April 2012): 413–23. https://doi.org/10.1080/01449291003793793.

259. Yadollahi, Elmira, Wafa Johal, Ana Paiva, and Pierre Dillenbourg. “When Deictic Gestures in a Robot Can Harm Child-Robot Collaboration.” In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 195–206. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202743.

260. Yarosh, Svetlana. “Designing Technology to Empower Children to Communicate with Non-Residential Parents.” *International Journal of Child-Computer Interaction* 3–4 (January 2015): 1–13. https://doi.org/10.1016/j.ijcci.2015.09.001.

261. Yarosh, Svetlana, Stephen Cuzzort, Hendrik Müller, and Gregory D. Abowd. “Developing a Media Space for Remote Synchronous Parent-Child Interaction.” In *Proceedings of the 8th International Conference on Interaction Design and Children - IDC ’09*, 97. Como, Italy: ACM Press, 2009. https://doi.org/10.1145/1551788.1551806.

262. Yarosh, Svetlana, Stryker Thompson, Kathleen Watson, Alice Chase, Ashwin Senthilkumar, Ye Yuan, and A. J. Bernheim Brush. “Children Asking Questions: Speech Interface Reformulations and Personification Preferences.” In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 300–312. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202207.

263. Yip, Jason, June Ahn, Tamara Clegg, Elizabeth Bonsignore, Daniel Pauw, and Michael Gubbels. "'It Helped Me Do My Science.': A Case of Designing Social Media Technologies for Children in Science Learning." In *Proceedings of the 2014 Conference on Interaction Design and Children - IDC '14*, 155–64. Aarhus, Denmark: ACM Press, 2014. https://doi.org/10.1145/2593968.2593969.

264. Yip, Jason C., Tamara Clegg, June Ahn, Judith Odili Uchidiuno, Elizabeth Bonsignore, Austin Beck, Daniel Pauw, and Kelly Mills. "The Evolution of Engagements and Social Bonds During Child-Parent Co-Design." In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 3607–19. San Jose California USA: ACM, 2016. https://doi.org/10.1145/2858036.2858380.

265. Yip, Jason C., Kiley Sobel, Xin Gao, Allison Marie Hishikawa, Alexis Lim, Laura Meng, Romaine Flor Ofiana, Justin Park, and Alexis Hiniker. "Laughing Is Scary, but Farting Is Cute: A Conceptual Model of Children's Perspectives of Creepy Technologies." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–15. Glasgow, Scotland Uk: ACM Press, 2019. https://doi.org/10.1145/3290605.3300303.

266. Yip, Jason, Tamara Clegg, Elizabeth Bonsignore, Helene Gelderblom, Emily Rhodes, and Allison Druin. "Brownies or Bags-of-Stuff?: Domain Expertise in Cooperative Inquiry with Children." In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13*, 201–10. New York, New York: ACM Press, 2013. https://doi.org/10.1145/2485760.2485763.

267. Yoon, Sang Ho, Ansh Verma, Kylie Peppler, and Karthik Ramani. "HandiMate: Exploring a Modular Robotics Kit for Animating Crafted Toys." In *Proceedings of the 14th International Conference on Interaction Design and Children - IDC '15*, 11–20. Boston, Massachusetts: ACM Press, 2015. https://doi.org/10.1145/2771839.2771841.

268. Yung, Amanda K., Zhiyuan Li, and Daniel Ashbrook. "Printy3D: In-Situ Tangible Three-Dimensional Design for Augmented Fabrication." In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 181–94. Trondheim Norway: ACM, 2018. https://doi.org/10.1145/3202185.3202751.

269. Zaman, Bieke, and Vero Vanden Abeele. "Laddering with Young Children in User EXperience Evaluations: Theoretical Groundings and a Practical Case." In *Proceedings of the 9th International Conference on Interaction Design and Children - IDC '10*, 156. Barcelona, Spain: ACM Press, 2010. https://doi.org/10.1145/1810543.1810561.

270. Zamuner, Tania S., Lara Kilbertus, and M. Weinhold. “Game-Influenced Methodology: Addressing Child Data Attrition in Language Development Research.” *International Journal of Child-Computer Interaction* 14 (October 2017): 15–22. https://doi.org/10.1016/j.ijcci.2017.09.001.

271. Zhou, Yinsheng, Graham Percival, Xinxi Wang, Ye Wang, and Shengdong Zhao. “MOGCLASS: Evaluation of a Collaborative System of Mobile Devices for Classroom Music Education of Young Children.” In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI ’11*, 523. Vancouver, BC, Canada: ACM Press, 2011. https://doi.org/10.1145/1978942.1979016.

272. Zimmermann-Niefield, Abigail, Makenna Turner, Bridget Murphy, Shaun K. Kane, and R. Benjamin Shapiro. “Youth Learning Machine Learning through Building Models of Athletic Moves.” In *Proceedings of the Interaction Design and Children on ZZZ - IDC ’19*, 121–32. Boise, ID, USA: ACM Press, 2019. https://doi.org/10.1145/3311927.3323139.

Chapter 3:

Make Me Smile: Exploring Cognitive Thought Processes in Pictorial Scale Designs for Children

1. Abstract

While self-reported accounts of experience from children are essential in designing and evaluating the technologies they interact with, scale design, particularly with young children, in the field of Child-Computer Interaction (CCI) remains underresearched. As we demonstrate in a digital large-scale questionnaire study (*N* = 3263) containing three pictorial response scales (i.e., Nodding heads, Smileyometer, Thumbs-up) involving children aged 6 to 8, the quality of children's survey responses can be compromised by including extreme responding and nonresponse. Therefore, we started to investigate the cognitive thought processes underlying children's responses with a cognitive interview study (*N* = 25). We show (a) how the children comprehend the scales and retrieve the information from memory, (b) how children judge and communicate their final answers, (c) how the scales are rated in terms of their preference, ease of use, and fun by the children, and (d) how the scales can be improved to safeguard potential biases in children's responses. The two primary disruptive outcomes of the survey responses, nonresponse and extreme response, might be attributed to the following underlying factors: satisficing, ambiguity in item and response formulation, labeling, and number of response options. Our results point forward critical aspects of scale design for children and discuss how reducing complexity and increasing fun in scale designs could be two essential ways to cope with biases in children's responses. Following our conclusion and design recommendations, future research on pictorial scale design for children has to carefully investigate the underlying variables to draw scientifically rigorous conclusions and generalize our findings.

Keywords: Child-Computer Interaction, Survey, Pictorial scale, Smileyometer, Thumbs-up

2. Introduction

Several interdisciplinary fields like Child-Computer Interaction (CCI) began to grow steadily (Hourcade, 2015) and along with it, the need to develop methods to design and evaluate how children interact with technologies along all stages of the design process (Lehnert et al., 2022). The use of surveys as an evaluation method is popular in the field of Human-Computer Interaction (HCI), and they have been widely used to collect opinions and data from individuals (Read, 2008); they are easy to administer and require a low effort from experimenters, and they are especially useful in domains where deeper forms of engagement are not possible (Lallemand & Gronier, 2018). However, surveying does pose distinctive methodological problems that could impinge on the quality of data (Scott, 2000). For example, survey questions can be misinterpreted by respondents, especially children, whose cognitive abilities (i.e., literacy and language use) are still maturing, which makes them more receptive to biases (Borgers et al., 2000). It is thus important to investigate how children comprehend and respond to survey questions to increase the validity and reliability of our instruments.

Research concerning the question-answer process in surveys with children is very rare (Borgers et al., 2000). Previous research has mainly focused on improving question formulation to increase validity, based on generic rules for questionnaire construction for adults (i.e., simple words, avoid ambiguity, one question at a time) (Alwin & Krosnick, 1991; Borgers & Hox, 2001; Borgers et al., 2000). Hence, methodological expertise on how to conduct well-designed response scales taking children's cognitive capacity, motivation, or contextual factors into account remains scarce. Extreme answers (Read et al., 2002.; Zaman et al., 2012), nonresponse, random answers, low repeatability, and inconsistency are often reported (Borgers, 2003; Borgers et al., 2004; Borgers & Hox, 2001), especially with young school-aged children (Read & MacFarlane, 2006; Zaman et al., 2012).

To address this methodological gap, the current paper investigates the cognitive thought processes underlying children's survey responses to uncover potential sources of errors in design scales for children. We obtained the data for this analysis from an initial investigation involving a large-scale digital questionnaire. This questionnaire implied the presence of certain inherent biases in children's responses (i.e., on the extremely positive end of the scale). Consequently, we conducted a study employing a cognitive interview technique to investigate the connection between cognitive processes, children's motivation, and the scale's complexity level. From the results, we derive recommendations for future scale design with children to inspire and guide researchers in designing more valid and reliable scales.

The following section provides more background and underlying theories informing our methodological choice. Also, we summarize earlier research studies for the three scale formats and conclude with our research objectives. In section 3, we describe the digital questionnaire study and the cognitive interview study. In section 4, we provide a detailed analysis of the findings. Section 5 discusses the results before concluding on the primary outcomes in section 6. The last section details the selection and participation of children in our study.

3. Related work

3.1. Cognitive processes underlying children's responses

A common problem in questionnaires with children relates to having confidence that the children understood the question so that the correct answer can be elicited (Markopoulos et al., 2008). The study of Cognitive Aspects of Survey Methodology (CASM) focuses on respondents' thought processes to assess the validity and potential sources of error in survey responses (Jobe, 1991; Schwarz, 2007; Willis, 2005). This methodology is based on Tourangeau's four-stage model of response process (Kano et al., 2010) which is widely used in the HCI field, to explain how respondents in a survey process respond to questions. According to this model, a well-designed survey question will prompt a respondent through four stages, underlying specific cognitive processes:

- **Comprehension** – understanding the question and the information required;
- **Retrieval** (for factual questions) – remembering or calling to mind the appropriate information;
- **Judgment** (for factual questions) – the processing of recalled information to form judgements;
- **Selection** – choosing and communicating an answer. Two sets of processes have been proposed: translating a judgment into the scales provided, and revising the response based on factors such as consistency;

Survey response effects can arise at any stage of the response process and a respondent can decide to skip certain stages of the model (Kano et al., 2010). Previous research (de Leeuw, 2011) identified several age-specific biases as having an influence on the survey responses across the different response stages. For children in the middle childhood challenges have been reported in the *comprehension stage* with ambiguity or indirect questions as they are not skilled in thinking abstractly and reflecting on intangible concepts (Piaget, 1955); they interpret words literally. In the *retrieval stage*, questions' complexity can cause difficulties with children's memory skills. Within the *judgment* stage, certain types of judgments are hard to make for children, as they are developing their capacity to reason; unbalanced response options should be avoided. In the *selection* stage, children between 7 to 10 years are able to give informative responses only if provided with clear and concrete response options; for example, no ambiguity of response scales (e.g., vague quantifiers/unclear quantifiers as *sometimes* or *often) and* response options should offer neutral midpoints and contain completely labeled response options (except for pictorial response scales where labels are not necessary). Research has shown (Borgers et al., 2004) that the optimal number of response options in early middle childhood (7 to 10) equals 2 to 3 options, in later middle childhood and early adolescence 4 to 5 options (de Leeuw, 2011).

Disregarding respondent and question-response characteristics can lead to biased outcomes and non-response in surveys. Researchers (Borgers & Hox, 2001; Hall et al., 2016) compare item nonresponse or extreme responding to a way of satisficing, that is, when the respondent gives responses that appear acceptable without going through all the steps in Tourangeau's model. The opposite strategy, optimizing, occurs when the respondent carefully goes through all the cognitive steps necessary to answer a question. The latter generates more reliable responses than satisficing (Borgers & Hox, 2001) .

Previous research has shown that child respondents have a tendency to select a positive response option more frequently and that they are more inclined to select the higher end of a scale, than the lower end (Read, 2008; Read & MacFarlane, 2006). Research investigating potential reasons for this tendency has focused mainly on the children's age as a decisive factor (Read et al., 2002; Read, 2008; Read & MacFarlane, 2006). However, satisficing can be mitigated by different factors; the task's difficulty, the respondent's cognitive ability, and the motivation to perform the task (Borgers & Hox, 2001; Krosnick et al., 1996a).

One way of supporting the cognitive thought process of children's responses in surveys can be accomplished by limiting the verbal load of a question-response option. Therefore, survey design for children has recently focused on applying pictorial scale design to overcome reliability issues caused by maturing cognitive abilities.

3.2. Pictorial scale design in children's research

Pictorial scales are anticipated to enhance motivation, direct attention, and provoke interest (Haddad et al., 2012; Valla et al., 1994) due to their reduced cognitive load, enhanced comprehensibility, and heightened enjoyment (Ghiassi et al. 2011; Sauer et al. 2021). A pictorial scale utilizes visual components to communicate the significance of its pieces, eliminating the need to express emotions through language (Kunin, 1955) (p. 66); the images should speak for themselves (Borgers et al., 2000). Pictorial scales have proven especially useful for respondents with insufficient language and reading skills (Sauer et al., 2021), for example, in multilingual contexts or among younger age groups. A problem reported with pictorial scales is that they are aesthetically more appealing, which might lead children to choose the most pleasing response instead of the one that best reflects their opinion (Haddad et al., 2012; Reynolds & Johnson, 2011). Another drawback is that the graphical representation may not be intuitively understandable due to oversimplifications of the design (Sonderegger et al., 2016) or due to the rely on outdated design principles (Betella & Verschure, 2016). Hence, this ambiguity can lead to severe measurement error (Sauer et al., 2021). To overcome this limitation, verbal instructions or hints can be offered in addition to the images; however, they are often discussed to be counterproductive (Borgers, 2003; de Leeuw, 2011).

Two prominent and widely used pictorial scales in CCI are the Smileyometer (SOS) (Fig. 2) and the Thumbs-up scale (TUS) (Fig. 3). In the following section, we briefly review these two pictorial scales and include a third scale, the Nodding Heads (NHS) which is implemented in a survey from the governmental monitoring program in Luxembourg. We discuss empirical evidence for their psychometric quality from prior research.

Smileyometer scale

The Smileyometer scale (SOS) has been developed within the Fun Toolkit by Read (Read & MacFarlane, 2006), an instrument for gathering experiences of children about technology (Hanna et al., 1999). It can be used before or after children experience the technology, to measure their expectations or their experienced feelings. The SOS has been widely implemented in evaluating children's experiences in contexts, such as museum technology (van Dijk et al., 2012) or games (Zaman et al., 2012). The SOS is a 5-points hybrid Likert scale; it uses smiley faces combined with a text-based response set (see Figure 1). The scale ranges from "awful" to "brilliant", with an increase towards positive evaluation from left to right. Based on a co-design study with children aged 8 to 9, the neutral condition was removed (the straight-line mouth in the neutral condition, changed to a weak smile - Point-3) (Read, 2008). The SOS is presented to the children in a horizontal row with supporting words under the faces; children are asked to tick one face.

Figure 1. The Smileyometer scale (Read, 2008).

Studies validating the Smileyometer scale have often questioned the suitability of the scale for a younger age group (Read et al., 2002; Read, 2008; Read & MacFarlane, 2006; Zaman et al., 2012). Across several studies, authors found that younger children (8-10 vs. 9-10 (Read, 2008) and 7-9 vs. 12-13 (Read & MacFarlane, 2006)) scored higher on the Smileyometer scale than older ones. Thus, the age of children has been identified as a significant factor influencing children's responses. Other factors for the SOS than the respondent characteristics have been less documented so far (Borgers et al., 2000).

Thumbs-up scale

The Thumbs-up scale (TUS) (figure 2) is a 5-point hybrid Likert scale that intends to measure the perceived skill level of the respondent in a particular task (e.g., keyboard typing skill) (Kano et al., 2010). The TUS combines images of thumbs with a common text-based response set. The scale ranges from "poor" to "very good", with an increase in positive evaluations from left to right and a red/green color scheme.

Figure 2. An example of the Thumbs-up scale (TUS), retrieved from (Kano et al., 2010).

The TUS can be applied to any question regarding skill perception, and it has been validated in one study with children 7 to 8 years and 9 to 10 years old (Kano et al., 2010). Findings support the high construct validity of the scale (based on high correlation scores with a word cloud scale TUS r= .89), indicating that TUS can be used with children as young as 7. However, the authors reported very high scores on the positive end of the scale compared to the cloud scale, indicating that the answers had low sensitivity (Kano et al., 2010). The TUS has been implemented in a few recent studies, such as for the evaluation of music technology (Danso et al., 2021) or within a co-design method for large workshops (Read et al., 2022). For the TUS, to the best of our knowledge, no age-related differences have been reported yet, and other factors influencing the scale responses have also been less investigated to date.

Nodding Head Scale

The Nodding Head Scale (NHS) is a pictorial Likert scale featuring nodding and pictorial shaking heads, with stripes indicating a faster movement at the outer scale options. It has been designed with the intention to avoid polarity issues with items that inquire negative emotions (e.g., math anxiety). The NHS is used without additional labels and it has been implemented in Luxembourgish School Monitoring Programme (LUCET, 2023). It has been used in different point variants; a 2-point variant of the scale is used for the first grade (6 to 7 years) and a 4-point variant is used for grade 3 (8 to 9 years) and grade 5 (10 to 11 years). In the original scale, it was decided to use no neutral midpoint. The scale exists in black and white (see figure 3) and in color (see figure 5). In the younger grade, all teachers moderate through the entire questionnaire and the pupils receive a written example before starting the questionnaire.

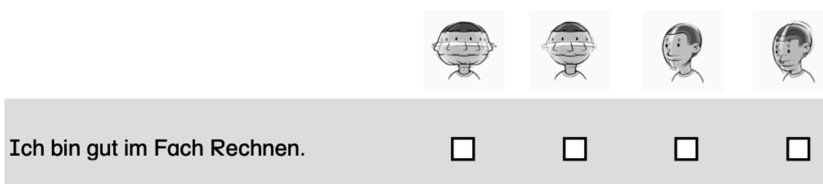


Figure 3. Sample item from a questionnaire intended to measure perceived mathematical skill for children aged 6 to 7, asking the children to rate the statement ‘I am good at mathematics’ on a 2-point scale.

There is no validation study for the NHS to date, and no effects of age are known, also with regards to the different scale answerpoints.

3.3. Scale evaluation

Throughout the reviewed papers (Kano et al., 2010; Read et al., 2002; Read, 2008; Read & MacFarlane, 2006; Zaman et al., 2012), the authors of the SOS and TUS focus on investigating the psychometric properties of the scales. However, as we learn from a recent review on pictorial scales in research and practice, there are more steps and methods which should be undertaken before evaluating the scale's validity and reliability (Sauer et al., 2021). Based on their review, the authors (Sauer et al., 2021) propose a three-phase approach to develop a pictorial scale: (1) item generation with methods such as brainstorming and rapid prototyping, (2) interpretation check with methods such as interviews and think-alouds, and (3) scale validation where a quantitative evaluation of the psychometric properties takes place. To our knowledge, neither one of the scales has included inspection methods to gain insights into user's mental models. For the development of the SOS, children were involved in a co-design study, which resulted in a slight difference in the design of the faces (Read, 2008). We would classify this kind of research in the item generation. To our knowledge, there are no user studies for any of the three selected scales that would address the mental models of the children and thus phase two of the approach. Therefore, we identified the need for a qualitative evaluation to understand why a particular scale might not be suitable for a particular group of children.

3.4. Research Objectives

As the selected scales (i.e., NHS, SOS, TUS) had already been published, and some had been used in several case studies (e.g., (Kano et al., 2010; Read, 2008), we trusted their validity and started with the implementation of our initial digital questionnaire. Additionally, our partners' interests and the yearly administration of e-assessments guided our choice to prioritize and begin with introducing the questionnaire in the field before launching the cognitive interview study.

However, due to the inherent biases (i.e., extremely positive responses) that we discovered during the analysis of the results of the digital questionnaire, we were forced to conduct a cognitive interview study to explain our quantitative results. We followed an explanatory sequential design (Creswell & Creswell, 2018), in which the researcher first does quantitative research, then looks at the results and uses qualitative research to go into more detail about what the results mean. This design may be counterintuitive because cognitive interviews or walkthroughs are usually done first so the researcher can address the changes before the quantitative study is conducted (i.e., exploratory sequential design (Creswell & Creswell, 2018)). However, due to the reasons mentioned earlier (i.e., previous implementation and partners' interest), we followed an explanatory sequential study design (Creswell & Creswell, 2018).

Consequently, with a cognitive interview study, we assessed the three scales more thoroughly to understand the thought process in survey responses with children. We focused on children between 6 and 8 years old as this was the age group where our questionnaire had been implemented.

We formulated the following research question: *How effective are the Smileyometer, Thumbs-up, and Nodding Head scales in supporting the thought process in survey responses with children between 6 and 8 years?* This general research question is broken down into four specific questions:

- RQ1: How do children comprehend the questions and retrieve the according information from memory to answer the question?
- RQ2: How do children judge and communicate their answers on these scales?
- RQ3: How are the scales rated in terms of their preference, ease of use, and fun by the children?
- RQ4: What are children's ideas for the (re)design of the scale?

The *cognitive interview* study investigates the four-stage model of thought process in-depth and helps to discover sources of confusion and misunderstanding to improve instrument design. We provide qualitative data to investigate RQ1 and RQ2. The first two stages of thought process (RQ1) and the third and fourth stage (RQ2) are taken together because these processes are often experienced as fluent. To complement the item response process, we have to include the survey question in our investigation, although the current paper focuses on the evaluation of scales.

Moreover, we investigated the scales' preferences, ease of use, and experienced fun (RQ3) in the cognitive interviews. We investigated these components as they are identified as influence factors on the satisficing strategies (Krosnick, 1991; Krosnick et al., 1996).

Finally, we provide design ideas of the scales by letting children brainstorming in the cognitive interviews (RQ4). This part of our investigation is based on phase one the three-phase approach to scale validation (Sauer et al., 2021).

4. Method

4.1. Study 1: Digital Questionnaire

Participants

Before the initial digital questionnaire study, we ran a pilot study during a large-scale tablet-based assessment (TBA) in two French schools. We implemented a paper-based version of our questionnaire with 73 pupils (i.e., *N* = 43 boys and *N* = 36 girls between 6 and 8 years). For the pilot study of the questionnaire, we received written consent from the parents for the participation of their child. The children received a certificate of participation to express our gratitude.

For the digital questionnaire, we collected data from 3261 pupils from 78 public French schools' CP (1st grade in US) (44.46 %) and CE1 (2nd grade in US) (55.54 %) grades, which correspond to age ranges 6-7 and 7-8 years respectively. For the CE1, we received data from *N* = 2,188, from which we removed 187 cases due to non-assent and 190 cases due to missing assent data. For CP, we received *N* = 2,187 and removed two cases due to duplicates, 329 cases due to non-assent, and 406 cases due to missing assent data. The final valid sample is *N* = 1,811 for CE1 and *N* = 1,450 for CP. For the digital questionnaire, the legal representative of the child gave consent to the Tablet-based assessments (TBA) conducted by the French Ministry of education, and our questionnaire was implemented as part of that study. We added an additional assent form into the TBA for the children: after a short introduction of the study purpose, the pupils had to state if they wanted to participate in the study by answering *yes* or *no* on a Nodding heads scale. As pointed out in detail above, we removed a total of 1112 participants from our data collection due to non-assent or missing assent. In addition, the study has been approved by the ethics board of the University of Luxembourg.

Material

The questionnaire was administered in the context of a TBA in France ("évaluation numérique socle"). The questionnaire content was chosen to address children's literacy with electronic devices (NHS) (LUCET, 2023), the perceived skill level of the respondent with a tablet (TUS) (Kano et al., 2010), and the experienced feeling (SOS) (Read et al., 2002; Zaman et al., 2012).

The outcomes from the pilot observation served the final design of the digital questionnaire, as outlined in table 1 and illustrated in figure 5. We made the following changes to the final questionnaire design based on the observation during the pilot study:

- The complete question and response options with all labels were added for each question (i.e., we replaced the sentence completion by full sentences)
- We splitted the question regarding the children's literacy with electronic devices (i.e., computer, tablet, smartphone) into separate questions;
- We asked for the children's ownership and their frequency of use for all the devices in a separate question;
- Instead of using *yes* and *no* (without pictorial response option) we introduced the NHS with accordingly two (e.g., question 2) and three response options (e.g., question 3);
- We standardized the voice input (i.e., a female voice-over was used to explain the questionnaire and study very short and then read aloud all question and answer options);
- The children could replay the voice-over of all questions and response options as often as they wished;

For the digital questionnaire implementation (Study 1), we conducted a pre- and post-measurement of the TUS and SOS similarly to the authors implementation (Kano et al., 2010; Read et al., 2002). Therefore, we decided on the following order of questions (see table 1) to connect the pre- and post-measurements of question 9 and 11 and question 10 and 12.

The complete questionnaire was read aloud by a female voice-over (see Figure 5 for an example of the questionnaire).

Figure 5. The Nodding Head scale implemented into the digital questionnaire design.

Table 1. The questionnaire description, with all answer and response options. In the Study 1, the questionnaire is divided into pre-questionnaire (before the TBA assessment) and post-questionnaire (after the TBA assessment). In Study 2, we skipped the post-questionnaire because the children were not exposed to an e-assessment.

#	Question	Scale	Response options
	Pre-questionnaire (Study 1 & 2)		
1	Do you want to answer our questions?	Nodding Heads	Yes / No
2	Do you have a computer at home?	Nodding Heads	Yes / No
3	Do you use that computer at home?	Nodding Heads	Yes, very often / Yes, sometimes / No, never
4	Do you have a tablet at home?	Nodding Heads	Yes / No
5	Do you use that tablet at home?	Nodding Heads	Yes, very often / Yes, sometimes / No, never
6	Do you have a smartphone at home?	Nodding Heads	Yes / No
7	Do you use that smartphone at home?	Nodding Heads	Yes, very often / Yes, sometimes / No, never
8	Do you use a tablet or computer for doing your homework?	Nodding Heads	Yes / No
9	How good are you at using a tablet?	Thumbs-up	Very difficult / somewhat difficult / somewhat easy / easy / very easy
10	How are you feeling right now?	Smileyometer	Very bad / not very good / good / really good / very very good

	Post-questionnaire (Study 1)		
11	How good are you at using a tablet?	Thumbs-up	Very difficult / somewhat difficult / somewhat easy / easy / very easy
12	How are you feeling right now?	Smileyometer	Very bad / not very good / good / really good / very very good

Procedure

We collected large-scale data with a digital version of the questionnaire that was integrated into a TBA application, randomly containing either a French or a Mathematics module. In figure 4, we illustrate the procedure of the field study.

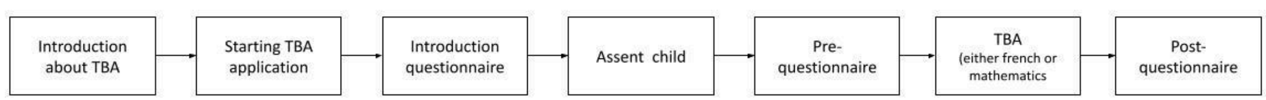


Figure 4. Procedure of the digital questionnaire during the Technology-Based Assessment (TBA).

Data analysis

For the large-scale study, we received anonymized .csv files from the French Evaluation, Forecasting and Performance Department (DEPP) with the outcomes from our questionnaire. After cleaning the data files, we transferred the data output to IBM SPSS Statistics V25 for further analysis.

We followed a quantitative evaluation method similar to (Hall et al., 2016). We looked into the following measurements:

- **Completion rates:** this assesses the percentage of children who completed the rating scales. Low completion rates can indicate a lack of motivation or misunderstanding of the question-answer process;

- **Individual Variance:** this identifies the variance within an individual's responses. High Variance (e.g., using the whole scale across different moments of measurement) can indicate that the scale provides children with a method that supports them in making judgments;
- **Sample Variance:** this assesses the variance between participants, determining if, within the whole sample, the entire scale within one question (e.g., for the Smileyometer, all 5 Likert-points) has been used. Higher variance (i.e., usage of the entire scale) across different participants would indicate that the scale supported them in making judgments;

4.1. Study 2: Cognitive Interviews

Participants

25 children (11 boys and 14 girls) with a mean age of 6.83 years old participated in the cognitive interviews. Most children were between 6 and 8 years old, with two exceptions (one child was almost six, and one just became nine years old). For the cognitive interviews, we received written consent from the parents by mail. The children received fair compensation and a certificate of participation to express our gratitude. In addition, the study has been approved by the ethics board of the University of Luxembourg.

Procedure

We conducted the cognitive interviews online via a video conferencing tool (i.e., Microsoft Teams), together with the children's legal representative. We used a combination of think-aloud and verbal probing techniques, similar to (Knafl et al., 2007). Children were not exposed to any assessment situation for this study as we needed all additional time and cognitive resources (e.g., for asking in-depth questions).

The procedure of the cognitive interviews foresees two consecutive phases. The first phase addresses comprehension of the questionnaire and retrieval of related information from memory (RQ1 and RQ2), while the second phase of the interview aims to compare the scales and collect children's suggestions for scale improvement (RQ3 and RQ4) (see fig. 6).

After an introduction, the moderator walked through all questions as described in table 1. After each question, the moderator conducted a verbal probing technique by asking the child the following questions, repeated for all items contained in the scale (phase 1):

- Why did you choose this answer?
- How easy or difficult was it to understand that question?
- How easy or difficult was it to answer that question?
- What do you think about the pictures (nodding heads, smiley, and thumbs-up scale)?

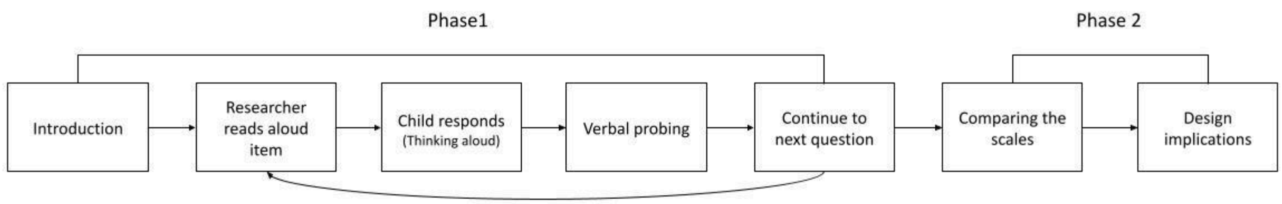


Figure 6. The procedure of the cognitive interview study.

Once all the TBA questions were answered, the researcher asked the following questions to the child (phase 2):

- Why do you think this scale is most fun/least fun?
- Why do you think this is easiest to use/most complicated?
- Which scale do you think other children would prefer most?
- What scale did you prefer? (we showed a picture with all scales and let them sort the scales)
- Why do you prefer this scale?
- How can we make the scale better? (we showed a text beneath each scale and asked the children to come up with a different visual representation)

As three children showed signs of tiredness, we skipped the last questions for them (i.e., “how can we make the scale better”). We finalized each interview by thanking the child and adult and sending out the voucher and certificate immediately after.

Material

For the cognitive interview study, we implemented only the pre-questionnaire (see table 1). The post-questionnaire was skipped because we did not confront the children with a TBA.

Data analysis

During the interviews, another researcher took notes. After the study, the moderator and the researcher completed the transcript based on the video recordings. Both then coded all transcripts from every participant in Microsoft Excel individually. As we started with an open coding approach (deductive coding), we met again to discuss and agree on the formulation of the codebook, resulting in a consensual codebook (see appendix). Then both authors coded the transcripts of all participants again individually. For this last round of coding, the interrater agreement was calculated in IBM SPSS Statistics V25 using correlation analysis, Kappa = 0.869 with $p < 0.000$.

5. Results

5.1. Study 1

In study 1, we followed a quantitative evaluation method. We used three measures, similar to Hall et al. (2016), that is, completion rates, individual and sample variance.

Completion Rates

We see that the NHS has the highest percentage of nonresponse for most cases, except for the CE1 grade TUS (see table 2). For the SOS, we observed no huge difference in nonresponse across both grades and a higher completion as for the NHS and TUS CE1. We observed a significant difference in completion between the grades for the TUS, with a high percentage of nonresponse for the CE1 grade (14.03 %) and a very low percentage for the CP (0.16 %).

Table 2. Summary of all completion rates and nonresponses for the three scales (Thumbs-up, Smileyometer, and Nodding Heads) across the two grades (CP and CE1).

Grade	Completion	TUS %	SOS %	NHS %	Total %
CP (6 to 7 years)	No	0.16	1.82	13.95	9.24
	Yes	99.84	98.18	86.05	90.76
CE1 (7 to 8 years)	No	14.03	1.24	11.71	10.23
	Yes	85.97	98.76	88.29	89.77

Furthermore, we see in table 3 that for the CP grade on the NHS scale the nonresponse percentage is higher on the 3-point Likert scale compared to the 2-point Likert scale. For the CE1 grade on the NHS, we almost did not observe a difference between the two Likert options.

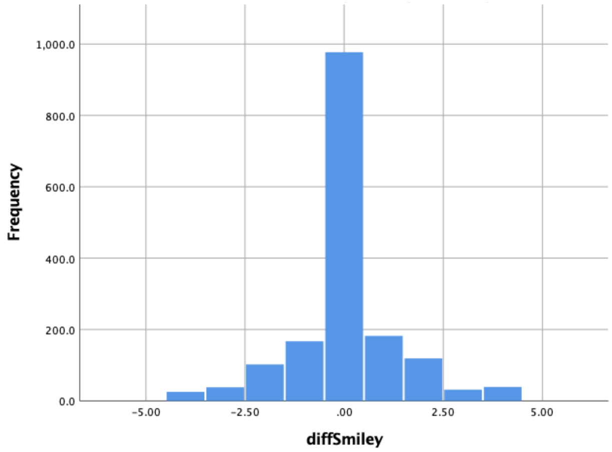
Table 3.

Non-completion rates across the two different response options of the NHS (2-point vs. 3-point Likert scale) for the two grades, CP and CE1.

Grade	N response options	NHS non-completion %
CP (6 to 7 years)	2	0.11
	3	32.41
CE1 (7 to 8 years)	2	12.29
	3	10.91

Individual variance

We investigated how the pupils changed their responses on the SOS (Q10 and Q12) and the TUS (Q09 and Q11). We used the following formula to calculate the individual difference in SPSS for the SOS: diffSmiley= post-smileyometer - pre-smileyometer and respectively, diffSmiley= post-smileyometer - pre-smileyometer for the TUS.



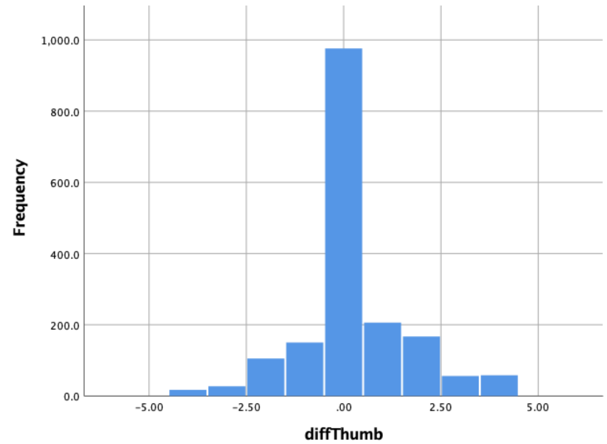


Figure 7. Bar chart illustrating the difference in responses on the pre- and post-SOS (left, figure 7a) and pre- and post-TUS (right, figure 7b) for the CE1 dataset.

In both figure 7a and 7b of the CE1 dataset, we see that most pupils did not change their responses on the scales; for SOS (*M* = .05, *SD* = 1.30829, *N* = 1,680), and for TUS (*M* = .2446, *SD* = 1.36592, *N* = 1,762). We observed similar findings in the CP dataset for the SOS (*M* = .2432, *SD* = 1.47841, *N* = 1,320) and TUS (*M* = .4661, *SD* = 144732, *N* = 1,401). Only CE1 data are shown as the CP visualization is very similar.

We further calculated Spearman's rank-order correlation to determine the relationship between the pupils' scores on the pre-SOS and post-SOS. There was a slight positive correlation which was statistically significant for the CE1 dataset (*rs*(1678) = .266, *p* = .000) and for the CP dataset (*rs*(1318) = .210, *p* = .000).

The correlation between the pre-and post-thumbs-up scale is close to 0 for the CE1 dataset (*rs*(1760) = .175, *p* = .000) and for CP (*rs*(1399) = .166, *p* = .000), indicating a non-monotonic relationship. No linear relationship is assumed.

As illustrated by the scatter plot presented in figure 8, pupils choose roughly the same responses on both scales before and after the assessment, with most responses on the extreme ends of the scales. Only CE1 data are shown as the CP visualization is very similar.

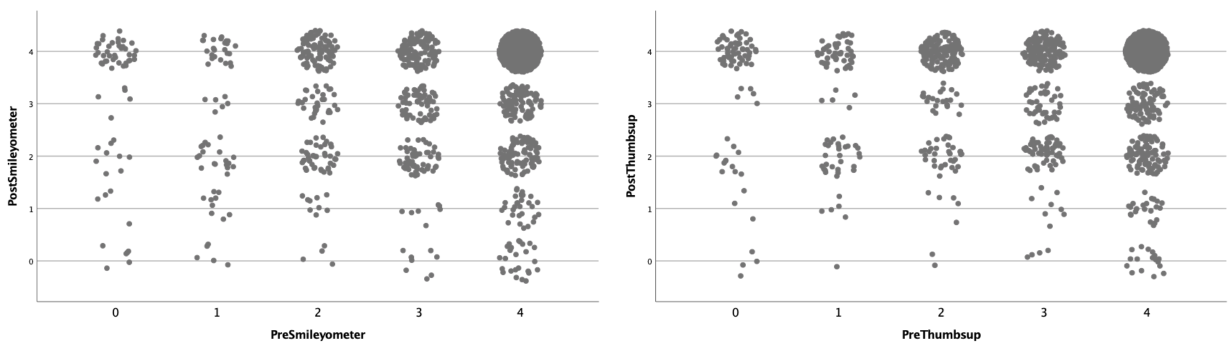


Figure 8. Scatter plot illustrating the correlation of the responses within the SOS (left) and TUS (right) for the CE1 dataset.

Sample Variance

We further investigate the variance between participants, determining if, within the whole sample, the entire scales had been used for each question.

We see a clear shift toward the extremely positive end for the SOS and TUS. A Mann-Whitney-U test showed that the scores of the CE1 grade were statistically significantly lower than for the CP grade on the pre-TUS (U = 1527846.50, Z = -4.724, $p < .001$) and the pre-SOS scale (U =1923984.00, Z = -2.193, p = .028), meaning that children from CP grade have an even stronger tendency toward the extreme positive end of the scales. The differences between grades in the post-TUS (U =1664668.00, Z = -1.078, $p < .281$) and post-SOS (U = 2029910.50, Z = -1.955, $p < .051$) scales were not significant .

Also, between the scales, we can observe a relationship between the pre-SOS and pre-TUS, meaning that the children who choose, for example, 3 on the pre-TUS also choose 3 on the pre-SOS. We run a Spearman's rank-order correlation to determine the relationship between the pupils' scores on the pre-SOS and pre-TUS. There was a positive correlation found, which was statistically significant for CE1 (rs(1691) = .601, p = .000) and for CP (rs(1315) = .529, p = .000).

We ran a Spearman's rank-order correlation to determine the relationship between the pupils' scores on the post-SOS and post-TUS. There was a positive correlation found, which was statistically significant for CE1 (rs(1356) = .391, p = .000).

5.2. Study 2

In this section, we summarize the findings from the cognitive interview study. We refer to the appendix, for a complete overview of all codes coming forth from the analysis of the three scales, including descriptions and examples along the different stages of the question-response model of thought.

Nodding Head Scale

Comprehension and retrieval (RQ1)

Regarding the comprehension of the questions for the NHS, we observed differences among the children. Some children could explain the word *smartphone* to us, and others did not know for sure what it was and what you could do with it ("It's a phone that takes 3D pictures" - P2). One child also experienced problems telling us if there was a tablet in her house because she was not sure if it was hers or the parent's tablet and hence, what she should report. One participant also got confused with the question if there was "one" computer at his house, because there were several computers. In total, we counted three children who expressed their understanding of the question-response options for the NHS and three children, who did not understand the meaning in the question-answer option.

Judgment and response (RQ2)

We observed that children generally understood the NHS with ease (N = 3) and some mimicked the movement of the head compared to two children who did not understand the scale. However, two children also took some cognitive effort to identify the movement of the head (e.g., "I couldn't see clearly what the head was doing" - P9). One child was also telling us that they found the movement of the Nodding Heads funny, whereas another one did not like the design (e.g., "it's a bit weird the heads like there are two guys with a stripe" -P8). Referring to the middle point of the 3-point NHS, the children have no problem differentiating it from the other points on the scale (e.g., "the *yes very often option,* he does move a lot with the head and one can see that well with all the lines (...) with the third one, *no never*, we see that he moves his head to the other side, not up and down, but to the side" -P1). Moreover, most children understood the difference between the *very often* and the *sometimes* answer (e.g., "very often means the whole day"..." sometimes means that I am not using it every day"- P4). One child also mentioned that the words describing the labels were useful to better understand the meaning of the scale. However, some children (*N* = 7) used alternative words to describe how often they used a tablet, phone, or computer. Also, three children mentioned that they would like additional response options like almost never. Interestingly, we also observed that children formed a judgment around their answers (e.g., "it means that he looks at the screen too much, that he is absorbed by the screens and he cannot let go of the tablet" - P5 and “I never use a phone because I don’t have the age for it”- P9). Some children also relied on parents' activities with the devices to respond to the question, for example, P14 said *yes, very often,* and after counseling with the mother, he corrected himself, saying “no, no zero!”. We also observed unexplained inconsistencies in the responses children were giving, for example, by saying *not so often* spontaneously and then choosing *no never* (P2) or saying “hyper rarely” and finally choosing *sometimes* (P23). We also discovered some inconsistency across the responses, for example, P12 said he used the tablet *very often,* whereas he did not have a tablet at home.

Thumbs-up Scale

Comprehension and retrieval (RQ1)

We found different perspectives on how children comprehend the question for rating their skill with using the tablet. They primarily relied on, what we call the *activity level* (*N* = 12), meaning that for estimating their capacity, they referred to an activity with the tablet, like watching videos or playing games. Another group of children (*N* = 6) relied on features regarding the *ease of interaction* with the tablet (e.g., "(...) you just click on things, move with your finger" - P9). Four children also approached this question by relying on their *skill level* with the tablet (e.g., “I know how to do everything on the tablet, I know how to use it well” -P6). Two children even went further in terms of skill level and compared their skills to others (i.e., “I am better at using the tablet than my grandmother” - P1). Other descriptions referred to *technical issues* (e.g., connections, bugs) (*N* = 1), or *frequency of use* (*N* = 2). One child had high expectations and compared themselves to others (e.g., “...sometimes I am even better than my mother…” - P12).

Judgment and response (RQ2)

A positive point mentioned by three children was the *familiarity* with the thumbs. One child particularly mentioned using the thumbs at school to indicate if she likes or does not like something. Another child was rather concerned with the *negative meaning* of the thumbs down (e.g., "because you put your fingers down, it means it's not good, and when you say not good, it's a little sad" - P14). Similarly to the NHS, one child mentioned that the scale was easy to mimic with the thumbs up and down (i.e., “because you just had to make the gestures"- P22).

Moreover, one child expressed a *negative association* with the colors used in the TUS (i.e., “the hands, I didn't like them at all since they look like the hands of ogres and witches with red and green”- P13). Contrarily, one child further stated that the color actually helped him to structure the scale responses into meaningful chunks (i.e., “red is not good, green means good” - P9). Yet another child indicated to not understand how the *colors are distributed* (e.g., “there are three green and two red thumbs, I do not understand that” - P18). Hence, some children felt overwhelmed by the *amount of the scale points* (i.e., “it was a little hard to answer and say it's good, it's not good, not very good” - P7) while one child did like to have plenty of options to choose from (i.e., “there are more answer options to choose from” - P18). Another child particularly mentioned missing the right words in the scale to describe what he wanted to answer (i.e., “I'm having a little trouble expressing myself when you asked me that question because I actually didn't directly find the words” - P4). One child also indicated that all the options made it hard for her to concentrate and one boy said that he has trouble “remembering too much about how to use the tablet well” (P5) and therefore found it difficult to answer.

Smileyometer scale

Comprehension and retrieval (RQ1)

Almost all children immediately understood the question of how they were feeling, and they relied on different perspectives to answer it. First and foremost, children (*N* = 4) talked about an *external event*, influencing their mood and, therefore, their rating on the Smileyometer scale. Two children explicitly mentioned an event from the past that had influenced their feelings (e.g., “because I was at school today and I don’t really like school” - P19), and one reported on a future event as being responsible for the actual feeling (e.g., "because on Thursday I'm going to the park, I'm going to have a picnic in the zoo”' - P25). Some children also talked about the excitement to participate in our study as the reason to feel good (e.g., “because I’m happy to meet you” - P1) (*N* = 6). Few children (*N* = 2) also reported *internal feelings* as influencing their rating on the Smileyometer scale (e.g., “I feel just very happy today” - P8). Negative emotions in children concerned rather some physical aspects (e.g., “I have a little headache…” - P13) (*N* = 2).

Judgment and response (RQ2)

Four children could easily identify themselves with the smileys to *express their mood* (e.g., "because I understand the emotion" - P16). Only one child said that they would not understand the *meaning of the smileys* without the description of the labels (e.g., "because if there wouldn't have been the writing, we didn't quite understand what it meant" - P12). One child said that they found the options overwhelming (i.e., "there are several good ones, I do not know what to choose" - P23), whereas another child only chose *very well* because there were not enough options. Regarding the design, children also expressed positive and negative points about the Smileyometer. Most prominently mentioned within the SOS was the *fun aspect* (e.g., "I like the smileys, they are fun" - P5). Another child also particularly mentioned liking the drawings of the smileys, opposite to yet another child who criticized the design (e.g., "the smiles are not well done (...) for very bad and really good, they are a little stuck to the chin" - P13). Similar to the Thumbs-up scale, a child also expressed the *similarity aspect* as positive about the scale (e.g., "I use the smileys a lot on Whatsapp" - P19) and that it was, therefore, easy to answer the question. Interestingly, we found a child arguing that the extreme answer was too strong to express her emotion ("I do not select *very well* because it's too strong" - P5). In general, children who selected *very well* had difficulties justifying why they had chosen this answer.

Preferences for the scales (RQ3)

To answer RQ3, we asked all children in the last part of our interview to rate the scales on the aspects fun, easiness, and preference. In 4 cases, a child did not respond to our question. Table 4 summarizes the findings.

Table 4. The ratings (we gave 1 point for the chosen option of the child) for characteristics of fun, easiness, and preference across the three scales.

Characteristic	Nodding Heads	Thumbs-up	Smileyometer	Total
The most fun	5	8	12	25
The least fun	11	7	5	23
The easiest	7	6	11	24
The least easy	12	8	4	24
Other children like most	3	2	14	19
I like most	3	5	17	25

From table 4, we can see that the Smileyometer scale is the scale reported as the most fun, most easy, and the preferred scale they would choose for other children and for themselves. In contrast, we see that the Nodding Heads scale is the least fun and the least easy to complete according to most children, and they think other children would not like the scale. The Thumbs-up scale has been reported primarily in between the two other scales and as the least preferred choice for other children.

Ideas for improving the scales (RQ4)

In general, a lot of children were recalling the scales they had seen previously and created combinations of the scales (e.g., “a head that smiles with a tooth" -P1). For the SOS, three children added an emotional layer to the scale (e.g.,”there are tears that fall from the hair” - P1). We also observed that children got inspired by the emoticon design to express certain scale responses (e.g., “a smile in the shape of a wave” - P1 or “crosses over the eyes” - P13). Other children got inspired by the use of the colors of the Thumbs-up scale to indicate bad and good accordingly (e.g., “a sad smiley in red and a smiley with a big smile to the ears in green” - P19). Regarding the frequency indications within the Nodding Heads scale (Q3, 5, 7), children found it more difficult to come up with representations for *very, often/sometimes/never*. One of the few suggestions we received were, for example, “a cross in front of my hand” (P2) or “someone who does nothing, with arms down” (P7). One child came up with contradictions to express the scale (i.e., “rain for bad and sun for feeling good” - P7). Last, we also heard many creative suggestions that were not directly linked to the response options, such as “a king, a dancing monkey or a french flag” (P14).

6. Discussion

Our analysis of the digital questionnaire (Study 1) showed that there are two major disruptive outcomes to children's survey responses. The first is **extreme responding**, where children answered on the very positive end of the scale as we observed in the individual and sample variance. The second is **nonresponse**, meaning that children did not respond to a survey question as we observed in the completion rates. We discuss the potential underlying factors responsible for these observations based on our cognitive interview study and literature.

6.1. Satisficing

One underlying reason responsible for the extremely positive answers and the nonresponse in children is related to the satisficing bias. Extreme responding happens, for instance, when respondents change their response before they communicate it (i.e., from the actual response to the response considered desired) due to influences of, for example, social desirability or situational adequacy (Kano et al., 2010). In our cognitive interviews some children judged the answers in the NHS, for example, "I don't use a smartphone because I don't have the age for it" (P9). In this particular case, we discovered later that the participant actually uses the phone of his father. However, he decided to reply with *no, never* to the question because it is what he thought is expected from him (i.e., social desirability). Additionally, the various inconsistencies in children's answers in our cognitive interviews (e.g., first choice is *sometimes* and then *no, never)* could indicate the use of satisficing strategies and cause extreme responding.

In our large-scale study, the digital questionnaire, we can exclude an effect of the experimenter as a cause for satisficing because we used a digital moderator for the questionnaire (i.e., female voice-over of the complete questionnaire and instructions). However, we cannot exclude the possibility of social desirability, as the questionnaire was administered in an assessment context where children might want to please their teachers, parents, or other authorities involved. Moreover, certain types of wordings in a question (i.e., "how good are you at…"-Q9 and 11) which relate to the motivations or beliefs of the participant may drive them to respond in an extreme manner (Krosnick et al., 1996a). Children also use satisficing strategies if they are not very interested in the topic or lose interest after a while due to their limited attention span (Borgers et al., 2004). This effect will be worse if the content is experienced as difficult (de Leeuw, 2011). We observed that the NHS, which was reported to be less fun and more complex compared to other scales in the cognitive interviews, was also the scale with the lowest completion rate in the digital questionnaire. We observed the reverse effect for the SOS, which had been reported as highly fun and less complex according to the children and containing the highest completion rates (see figure 9). This illustrates how satisficing strategies can cause nonresponse in children's answers. Similarly, other authors (Borgers, 2003; Borgers et al., 2000; Krosnick, 1991; Krosnick et al., 1996) have reported how the difficulty of the task (in fig. 11 referring to complexity), the motivation (in fig. 11 referring to fun) and the cognitive abilities of the respondent may cause satisficing strategies.

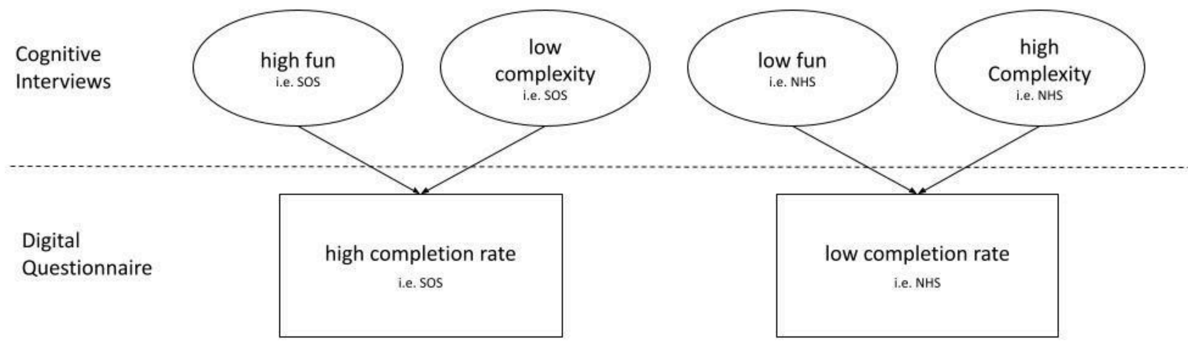


Figure 9. The figures illustrate a potential relationship between fun and complexity (as measured in the cognitive interviews) with completion rate (as measured in the digital questionnaire).

Regarding cognitive abilities, we found that for the younger age group (CP grade), the non-response percentage (in fig. 11 referring to the completion rate) is significantly higher on the 3-point Likert scale compared to the 2-point NHS. For the CE1 grade, we almost did not observe a difference between the two Likert options which could illustrate the relationship of the cognitive abilities and the complexity of a scale. Also, similar to previous research (Read, 2008; Read & McFarlane, 2006), we confirmed that the younger age group (CP grade) has an even stronger tendency toward the extreme positive end of the scales (i.e., SOS). We, therefore, assume that the cognitive abilities (i.e., age of the children) can have a mediating factor in the response output provided by the children.

6.2. Ambiguity of question-response options

Moreover, extreme responding and nonresponse might be caused by the question format itself due to a misunderstanding in the wording of the question (stage 1 of the response model of thought process) (Borgers & Hox, 2001). In line with (Krosnick et al., 1996a), the task's difficulty can affect the ease with which a survey question is understood and causes satisficing (Borgers & Hox, 2001). In our cognitive interviews, the TUS questions were interpreted very differently by the children, for example, some referred to the activity level (i.e., what they can do with the tablet) versus others who referred to the skill level (e.g., being better than someone else with using the tablet). The different ways of interpreting the TUS can cause invalid answers, and showing in, for example, the extreme responding. Moreover, the thumbs' images on the scale labels are not directly linked to the item's meaning (i.e., referring to the skill level), and it might therefore disturb the thought process (i.e., stage 3) (Tourangeau et al., 2000). The author refers at this juncture to a possible ambiguity of the item slowing down cognitive processes due to an overload in working memory (Tourangeau et al., 2000). To solve this, we could also provide pictorial content to convey the item's meanings to the children as this could, similarly to the rating scale, enhance the understanding. Therefore, it might be useful, especially for younger children, to have integrated entities (i.e., conveyor of item meaning and rating scale represent one entity) like in the SAM questionnaire (Bradley & Lang, 1994). In particular, in a scenario involving repeated measurements, such as the SOS and TUS in our example, utilizing a visual representation, such as a symbolic image, may facilitate more efficient and rapid information encoding (Bennett & Walters, 2001).

Furthermore, researchers have discovered notable variations in the dependability of survey results, particularly when children are asked to provide attitudinal judgments (Alwin & Krosnick, 1991). Researchers claim that, with attitude questions like the TUS and the SOS, respondents mostly form a judgment on the spot based on whatever relevant information is accessible then; and they do not retrieve the according information from memory as the researcher actually would like the respondent to do so (stage 2 is skipped) (Tourangeau et al., 2000). Younger children especially struggle with questions administering intangible and theoretical constructs (Borgers et al., 2004). As we have seen with the SOS question, we received manifold insights into why a child in a particular moment is feeling the way she/he did; from past external events (e.g., a bad day at school) to internal feelings (e.g., I have a headache). Therefore, we might wonder, how much those events actually relate to the experience with the tablet-based assessment, which we were aiming to measure as for most children, we did not observe a notable variance (pre-and post-scores on the TUS and SUS). Moreover, the sample variance in the large-scale study underlines this, as the correlation between the TUS and SUS scales could give us evidence that both scales are not measuring two distinct constructs. We, therefore, should question the effectiveness of the scales at this juncture to measure the user experience with the assessment system for those age groups as other authors similarly have in the past (Zaman et al., 2012).

Last, we observed that the TUS and SOS are familiar concepts for most children as they use the thumbs, for example, in school and the smileys on Whatsapp. Familiarity can enhance understanding (and i.e., lower complexity) in younger children as it eases stage 2 (Tourangeau et al., 2000), making it easier for children to retrieve information from memory. Younger children had lower missing rates for the TUS and SOS, which could also be due the familiarity with the scales.

6.3. Amount of Response options

Authors claim that a reason for extremely positive responses could be that the children are not provided with an **adequate set of responses,** thus failing to meet stage 4 of the response model (Hall et al., 2016). One way of addressing this could be to reduce the response options (de Leeuw, 2011). In our data, we saw that younger children especially benefit from the scale with lower response options, that is, the NHS. We assume that the older group invested more time in reading and evaluating the answer options. Younger children think mostly dichotomously, which was best represented in the NHS with the answer options *Yes* and *No* (Piaget, 1955). Consequently, we would be interested in validating if a 3-point SOS or TUS would make the scale more sensitive. This could be investigated in future studies evaluating the scale designs for children. Yet another approach could be to break down the standard Likert type response (agree strongly, agree, neither agree nor disagree, disagree, disagree strongly) into smaller pits by asking the child first if s(he) agrees or disagrees and then probing for strength of feeling (Borgers et al., 2000).

6.4. Scale Labels

As we learned from previous research, the scale labels (i.e., labels vs. partly labeled response scales) can also have an influence on the item nonresponse (Borgers & Hox, 2001). Children differ regarding their interpretation and ease of use for the different response options for the NHS. Some children found those labels under the pictures helpful, for others, it led to confusion. First and foremost, children give their own interpretation of the pictures, and if that's not matching with the labels, they end up confused and might not respond. The labels for the 3-point Likert scale in the NHS (i.e., sometimes, often, etc.) did not correspond with the images illustrating a shaking head. This might correspond with the very high nonresponse for the younger grade (more than 30%). Labels can add an additional cognitive burden to the thought process, and we, therefore, should carefully evaluate the advantage of adding them in addition to the pictorial response options. Especially, considering that other researchers found out that all age groups are equally affected by a more difficult question format, we should question the added value of response labels (Borgers, 2003).

6.5. Design Recommendations

Based on the results, we propose the following design recommendations to design scales supporting the response model of the thought process for children between 6 to 8 years:

- **Reconsider the use of attitudinal judgments:** it is hard for children to judge their feelings rationally. We have seen in our cognitive interview studies that the range of interpretation of attitudinal questions is quite diverse. Therefore, we recommend phrasing our research around the evaluation of tangible objects for children below the age of 9 years or to use dichotomous question-response alternatives.
- **Use stronger visual differentiation**: Children explicitly mentioned the different colors as helpful and attractive to them to respond to the TUS. The colors help structure and cluster the response options and let children encode the information faster and more efficiently. However, be aware of implicit biases introduced by colors (red-bad and green-good; use neutral colors alternatively).
- **Use what is known:** We favor what is already known to us. New things which fall entirely out of our mental model are considered less attractive and increase the cognitive load. Therefore, future scale design should investigate elements already part of the children's world and how these elements can be used to design surveys. The NHS has not been favored, having the highest complexity rating and having the lowest completion rates. We assume that the concept diverged too strongly from what children were familiar with (i.e., the smileys and thumbs).
- **Take out the labels:** Especially for children, whose reading skills are not fully developed yet, we do not need extra clarification of the response options by adding labels. This might form an additional cognitive burden on the thought process of children. If we decide to keep the labels, we should carefully check for their representation and interpretation along the spectrum of answers.
- **Use common entities:** to represent the item's meaning, it might support young children's response process if the question aligns with the response scale and is presented by pictorial content,

similarly as we have seen with the standardized questionnaire SAM. All pictorial item content should be carefully evaluated for their unambiguity.

- **Make Me Smile:** Scale design for children should be fun and make them smile as that could result in fewer nonresponses. We have seen that the SOS, which combined fun and low complexity, showed the highest completion rates in the questionnaire. Fun might have an effect on children's motivation to complete survey questions and can therefore reduce bias in survey design.

7. Limitations and future work

This research holds several limitations. First, the translation of the material into French and back to English for this paper – despite strict translation processes – could have changed the meaning of the scale responses and their underlying questions. Future studies have to investigate in more detail how language and translation of scale responses can affect the evaluations. Second, the qualitative data collected online made it necessary for some parents to assist their children during the video call. This might have biased the answers children gave (e.g., pleasing the parents in addition to the researcher). However, as this was a new situation for most children, we wanted to avoid them feeling nervous or anxious. Also, in the large-scale study, children may adopt a school attitude by answering our questionnaire, resulting in less nonresponse compared to questionnaires applied in other settings. Third, the selection of scales for the empirical evaluation has been based on research interest. We, therefore, avoid making claims or generalizations regarding these scales to another context or way of implementation. In order to make the scales comparable, we change critical aspects to the scales validity, for example, NHS has been used without additional labels in the original implementation and we have changed the amount of answer options. Therefore, the outcomes can not be generalized to other formats of scale implementation such as for the NHS in the school monitoring program.

Fourth, the scales are distributed over the questionnaire in a specific order and in different proportions - there may be a bias in this design which we have not further investigated. Previous split ballot study with school aged children (10 years and above) found that the children were prone to a primacy effect (Scott, 2000) so future studies could investigate this further. The study settings in the digital questionnaire closely resemble the situations encountered in ordinary life, preserving the environment's genuineness (Farrell & Fessenden, 2024) (e.g., enabling a comprehensive evaluation to investigate the behavior of variables in a real-world context). Therefore, our primary emphasis was documenting and studying existing phenomena rather than actively changing them. Field research helped us because it allowed us to apply our findings to real-world situations (i.e., our partners' large scale assessment) and covered a more natural surrounding (i.e., the classroom where the assessment is taking place). Nevertheless, with these advantages came the inability to control for certain factors (e.g., peer influence) and variables, making it difficult to draw scientifically rigorous conclusions and generalize our findings. Fifth, the scales were used in a digital version of the questionnaire; therefore, we do not generalize our findings to other formats, for example, paper-based questionnaire versions, because the way of interacting with the scale can affect the evaluation. Future studies could further investigate the effects of scale design on different formats and devices. In general, it is known that using computer-assisted questionnaires produces better-quality responses (De Leeuw et al., 2003) and this offers the possibility of standardizing the moderation of the questionnaire. Last, follow-up studies should investigate if new scale designs or changes to an existing scale like the SOS lead to better validity and reliability. Also, as pointed out earlier, it is important for very young children to identify with and be able to mimic a scale (i.e., embodied cognition). We could rethink traditional scale design, and let children respond to survey questions with their gestures and use appropriate software (e.g., face/body recognition) to track their responses. Exploring beyond conventional methods of surveying children (i.e., pen and paper or tablets), reveals several innovative alternatives that extend beyond the limitations of existing surveying technologies, for instance the use of tangible objects. These new tools might be more natural and put less cognitive load on the thought process, particularly beneficial for preschool-age children. However, assessing attitudinal and emotional aspects stays tricky, especially the younger the children are, but these psycho-motivational constructs are important facets of our research in the Child-Computer Interaction field.

8. Summary & Conclusion

This paper has empirically evaluated three pictorial scale designs, that is, Nodding Heads (NHS), Smileyometer (SOS), and Thumbs-up (TUS) with children between 6 and 8 years old. Our results highlighted different outcome biases of survey questions; extreme positive responding and nonresponse. We showed how important it is to design scales that are fun and, at the same time, less complex (i.e., fewer response options, no ambiguity) as this might influence children's responses. In our interview study, the NHS was evaluated as being the least fun scale and, similarly, the most complex one, and showing the highest nonresponse in the field study. Opposite observations were made for the SOS, as children rated the scale as highly funny and less complex compared to the other scales. This rating was in accordance with the observation in the digital questionnaire where the SOS showed overall higher completion rates. Future scale designers should investigate if improving fun and lowering the complexity of the scales can overcome bias in children's surveys, and more accurately investigate the interaction of different intervening variables to the scale designs. We derived design recommendations from our cognitive interview study guiding future scale design for young school-aged children.

9. References Chapter 3

Alwin, D. F., & Krosnick, J. A. (1991). The reliability of survey attitude measurement: The influence of question and respondent attributes. *Sociological Methods & Research*, *20*(1), 139–181.

Bennett, K. B., & Walters, B. (2001). Configural display design techniques considered at multiple levels of evaluation. *Human Factors*, *43*(3), 415–434.

Betella, A., & Verschure, P. F. (2016). The affective slider: A digital self-assessment scale for the measurement of human emotions. *PloS One*, *11*(2), e0148037.

Borgers, N. (2003). Response Quality in Survey Research with Children and Adolescents: The Effect of Labeled Response Options and Vague Quantifiers. *International Journal of Public Opinion Research*, *15*(1), 83–94. https://doi.org/10.1093/ijpor/15.1.83

Borgers, N., de Leeuw, E., & Hox, J. (2000). Children as Respondents in Survey Research: Cognitive Development and Response Quality 1. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, *66*(1), 60–75. https://doi.org/10.1177/075910630006600106

Borgers, N., & Hox, J. (2001). Item Nonresponse in Questionnaires with children. *Journal of Official Statistics*, *17*(2), 321–335.

Borgers, N., Hox, J., & Sikkel, D. (2004). Response Effects in Surveys on Children and Adolescents: The Effect of Number of Response Options, Negative Wording, and Neutral Mid-Point. *Quality & Quantity*, *38*(1), 17–33. https://doi.org/10.1023/B:QUQU.0000013236.29205.a6

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59.

Creswell, J. W., & Creswell, J. D. (2018). *Mixed methods procedures. Qualitative, quantitative, and mixed methods approaches* (Fifth edition). SAGE Publications, Inc.

Danso, A., Rousi, R., & Thompson, M. (2021). Novel and Experimental Music Technology Use in the Music Classroom: Learning Performance, Experience, and Concentrated Behavior. *Human Technology*, *17*(1), 81–112. https://doi.org/10.17011/ht/urn.202106223979

de Leeuw, E. (2011). *Improving data quality when surveying children and adolescents: Cognitive and social development and its role in questionnaire construction and pretesting*. Annual Meeting of the Academy of Finland: Research Programs Public Health Challenges and Health and Welfare of Children and Young People, 10-12.

De Leeuw, E., Hox, J., & Kef, S. (2003). Computer-assisted self-interviewing tailored for special populations and topics. *Field Methods*, *15*(3), 223–251.

Dillman, Don A. (2007). *Mail and Internet Surveys. The tailored design method.* (2nd ed.). John Wiley & Sons.

Druin, A. (Ed.). (1999). *The design of children's technology*. Morgan Kaufmann Publishers.

Farrell, S., & Fessenden, T. (2024, December 1). Field Studies [Nielsen Norman Group]. *Field Studies*. https://www.nngroup.com/articles/field-studies/

Fowler, F.J. (1995). *Improving Survey questions* (Vol. 38). Thousand Oaks: Sage.

Ghiassi, R., Murphy, K., Cummin, A. R., & Partridge, M. R. (2010). Developing a pictorial Epworth sleepiness scale. *Thorax*, thx-2010.

Haddad, S., King, S., Osmond, P., & Heidari, S. (2012). Questionnaire design to determine children's thermal sensation, preference and acceptability in the classroom. *Proceedings of the 28th International PLEA Conference on Sustainable Architecture+ Urban Design: Opportunities, Limits and Needs-Towards an Environmentally Responsible Architecture*.

Hall, L., Hume, C., & Tazzyman, S. (2016). Five degrees of happiness: Effective smiley face likert scales for evaluating with children. In *Proceedings of the 15th international conference on interaction design and children*, 311-321.

Hanna, E., Risden, K., Czerwinski, M., & Alexander, K. (1999). The role of usability research in designing children's computer products. In *The design of children's technology*, 4–26. Morgan Kaufmann.

Hourcade, J. P. (2015). *Child-Computer Interaction*. *Self, Iowa City, Iowa*.

Jobe, J. B. (1991). Cognition and survey measurement: History and overview. *Applied Cognitive Psychology*, *5*, 175–192.

Kano, A., Horton, M., & Read, J. C. (2010). Thumbs-up scale and frequency of use scale for use in self reporting of children's computer experience. *Proceedings of the 6th Nordic Conference on Human-Computer Interaction Extending Boundaries - NordiCHI '10*, 699-702. https://doi.org/10.1145/1868914.1869008

Knafl, K., Deatrick, J., Gallo, A., Holcombe, G., Bakitas, M., Dixon, J., & Grey, M. (2007). The analysis and interpretation of cognitive interviews for instrument development. *Research in Nursing & Health*, *30*(2), 224–234.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Appl. Cognit. Psychol.*, *5*(3), 213–236. https://doi.org/10.1002/acp.2350050305

Krosnick, J. A., Narayan, S., & Smith, W. R. (1996a). Satisficing in surveys: Initial evidence. *New Directions for Evaluation*, *1996*(70), 29–44. https://doi.org/10.1002/ev.1033

Krosnick, J. A., Narayan, S., & Smith, W. R. (1996b). Satisficing in surveys: Initial evidence. *New Directions for Evaluation*, *1996*(70), 29–44. https://doi.org/10.1002/ev.1033

Kunin, T. (1955). The construction of a new type of attitude measure 1. *Personnel Psychology*, *8*(1), 65–77.

Lallemand, C., & Gronier, Guillaume. (2018). *Méthodes de design UX. 30 méthodes fondamentales pour concevoir des expériences optimales.* (2nd ed.). Groupe Eyrolles.

Lehnert, F. K., Niess, J., Lallemand, C., Markopoulos, P., Fischbach, A., & Koenig, V. (2022). Child–Computer Interaction: From a systematic review towards an integrated understanding of interaction design methods for children. *International Journal of Child-Computer Interaction*, *32*, 100398. https://doi.org/10.1016/j.ijcci.2021.100398

LUCET. Retrieved on February 8, 2024. *Épreuves Standardisées*. https://epstan.lu/en/general-information/

Markopoulos, P., Read, J. C., MacFarlane, S., & Hoysniemi, J. (2008). *Evaluating Children's Interactive Products*. Morgan Kaufmann.

Pdl, A. (2018, January 22). Retrieved on February 8, 2024. *Not so innocent: Children and Fools Tell the Truth*.https://northartswest.wordpress.com/2018/01/22/not-so-innocent-children-and-fools-tell-the-truth/

Piaget, J. (1955). *The Construction of Reality in the Child*. NYBasic Books.

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*, 81-88. https://doi.org/10.1145/1139073.1139096

Read, J., Larusdottir, M. K., Islind, A. S., Sim, G., & Fitton, D. (2022). Tick Box Design: A bounded and packageable co-design method for large workshops. *International Journal of Child-Computer Interaction*, *33*, 100505. https://doi.org/10.1016/j.ijcci.2022.100505

Read, J. C., MacFarlane, S., & Casey, C. (2002). Endurability, engagement and expectations: Measuring children's fun. In *Interaction design and children*, 2(2002), 1-23. Eindhoven: Shaker Publishing.

Reynolds, L., & Johnson, R. (2011). Is a picture is worth a thousand words? Creating effective questionnaires with pictures. *Practical Assessment, Research, and Evaluation*, *16*(1), 8.

Sauer, J., Baumgartner, J., Frei, N., & Sonderegger, A. (2021). *Pictorial Scales in Research and Practice: A* *review*. *26*(2), 112–130. https://doi.org/10.1027/1016-9040/a000405

Schwarz, N. (2007). Cognitive aspects of survey methodology. *Applied Cognitive Psychology*, *21*, 277–287.

Scott, J. (2000). Children as Respondents: The Challenge for Quantitative Methods. In *Research with Children: Perspectives and Practices*, 98–119. Falmer Press.

Sonderegger, A., Heyden, K., Chavaillaz, A., & Sauer, J. (2016). *AniSAM & AniAvatar: Animated visualizations of affective states*, In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems,* 4828–4837.

Speake, J. (2015). *Oxford Dictionary of Proverbs* (6th ed.). Oxford University Press.

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive Processes Underlying Context Effects in Attitude Measurement. *Psychological Bulletin*, *103*(3), 299–314.

Tourangeau, R., Rips, J. L., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge University Press.

Valla, J. P., Bergeron, L., Bérubé, H., Gaudet, N., & St-Georges, M. (1994). A structured pictorial questionnaire to assess DSM-III-R-based diagnoses in children (6–11 years): development, validity, and reliability. *Journal of Abnormal Child Psychology*, *22*, 403-423.

van Dijk, E. M. A. G., Lingnau, A., & Kockelkorn, H. (2012). Measuring enjoyment of an interactive museum experience. *Proceedings of the 14th ACM International Conference on Multimodal Interaction - ICMI '12*, 249. https://doi.org/10.1145/2388676.2388728

Willis, G. B. (2005). *Cognitive Interviewing: A Tool for Improving Questionnaire Design.* Sage Publications.

Zaman, B., Abeele, V. V., & Grooff, D. D. (2012). Measuring product liking in preschool children: An evaluation of the Smileyometer and This or That method. *International Journal of Child-Computer Interaction*, *1*(2), 61–70.

10. Appendix

A. Overview of all codes for the Nodding Head Scale including descriptions and examples along the different stages of the question-response model of thought by Tourangeau.

Stage	Code	Description	Example Quotation
Comprehension /Retrieval	Clear wording in question	Understands the meaning of Question (Computer)	“We have a computer like this and a big one where you can watch movies on”
Comprehension /Retrieval	Clear wording in question	Understands the meaning of Question (Tablet)	“Yes, we have two tablets, one for me and one for Papa”
Comprehension /Retrieval	Clear wording in question	Understands the meaning of Question (Smartphone)	“A smartphone? A gsm that can do a lot of things”
Comprehension /Retrieval	Unclear wording in question	Does not understand the meaning of Question (Computer)	“I don’t know the difference between a computer and laptop”
Comprehension /Retrieval	Unclear wording in question	Does not understand the meaning of Question (Tablet)	“No, maybe yes, there is a tablet from me”
Comprehension /Retrieval	Unclear wording in question	Does not understand the meaning of Question (Smartphone)	“I don’t know what a smartphone is”
Judgment	Clear scale representation (design)	Understands the meaning of the Nodding Head Scale	“They (nodding heads) are well done, the first one says yes like that (imitating the movement) and the other no” "It’s funny how he moves the head" "Each time you have to take your time to do it (moving your head)"
Judgment	Unclear scale representation (design)	The interpretation of the design or some parts of the scale is not clear	"it's a little weird the heads, like there are two men with a stripe" “There was nothing funny in it. There was just a man who turned his head…and we're a bit used to seeing

			images of people moving"
Judgment/ Response	No explanation	No explanation or justification given	
Judgment/ Response	Clear scale representation (text)	Understands the meaning of the Answer option (Frequency)	"Because I don't do a lot of trainings where I use the tablet. Often would mean every day"
Judgment	Unclear representation (text)	Does not understand the meaning of the Answer option (Frequency)	"I don't understand the difference between often and sometimes"
Judgment/ Response	Inconsistent choice	Chooses a different answer in the scale than the spontaneous one	"Yes, always. I'm only allowed to use it if there is no school on Saturdays and Sundays"
Judgment/ Response	Missing scale option	The given answer is not in the scale	"Once a week"
Judgment	Response frequency / positive	The child likes that the scale has not so many answer options.	"…because there a less answer options"
Judgment	Response frequency / negative	The child has not enough answer options to express her/his opinion.	"Because sometimes you don't have the answer you found. And then you say to yourself hmm, do I take this, this, or that then."

B. Overview of all codes for the Thumbs-up Scale including descriptions and examples along the different stages of the question-response model of thought by Tourangeau.

Stage	Code	Description	Example quotation
Retrieval	Interaction	Rely on the ease of the (tactile) interaction	"it's because sometimes there are things you can't know if you're a child. (...) it's because there are little buttons that we sometimes don't know the meaning of."
Retrieval	Others	Rely on the ease / difficulty of use for others	"Because if something is not working my dad could fix it and he is better than my mother…sometimes I'm even better than my mother (...) But not better than dad"
Retrieval	Activity	Rely on the ease of the activity	"Sometimes I manage to do stuff on the tablet, games or something else, watch videos...."
Retrieval	Frequency of use / High	Rely on the frequency of the use	"because I use it often, (...) because I do it very often and I find it easy"
Retrieval	Skill oriented	Rely on the capacity (how to use)	"Because I have no trouble using it (...) For me that means that I know how to use it"
Retrieval	DiffTech	Rely on the technical issues (e.g., connection, bug..)	"because it's not hard but sometimes there are bugs (...) you just have to press things"
Retrieval	Frequency of use / Low	Rely on the low frequency of the use	"Because I'm not very much used to using a tablet"
Comprehension	Unclear wording in question	Does not clearly understand the meaning of question	"I didn't really understand the question"

Retrieval / Judgment	Color / positiv	Makes a positive statement regarding the color of the scale	"it shows, when it's red it's not good and when it's green it's good"
Retrieval	Familiarity	Recognises/related the thumbs-up scale from another context of use	“at school we do this to say if we are friends or if we are not friends”
Retrieval / Judgment	Mood	Relates the scale to the mood	"Because it's more like when we say we are very very well”
Judgment / Response	Answer options / positive	Does like the wide range of answer options	”because there are more answer options”
Judgment	Color / negativ	Does not understand the distribution of the colors	“I don't really understand because there are 3 green thumbs and two red thumbs”
Retrieval / Judgment	Unclear	The meaning of the scale is not clear	“it's harder to understand the meaning because we don't know what that (the thumbs) means”
Retrieval	Negative association	Has a negative association with the thumbs	“the hands, I didn't like them at all since they look like the hands of ogres and red and green witches"
Judgment / Response	Gestures / positiv	Likes that the scale expresses a gesture	“because you just had to make the gestures"
Judgment / Response	Answer options / negativ	Does not like the wide range of answer options	“it was a little hard to answer and say it's good, it's not good, not very good... It was hard to answer"

C. Overview of all codes for the Smileyometer Scale including descriptions and examples along the different stages of the question-response model of thought by Tourangeau.

Stage	Code	Description	Example quotation
Judgment	Participation Study	Relate their feeling to the participation in the study	"…and also because I like to do the quiz *(the interview)*"
Retrieval	External Event / positive	Rely on a positive event to express the feeling	"Because today everyone is nice to me..."
Retrieval	External Event / negative	Rely on a negative event	"because I was at school, I don't really like school"
Retrieval	Internal Pain / negative	Rely on a negative body sensation	"I have a little headache…"
Retrieval	Internal Feel / positive	Rely on a global positive mood without reason given	"because I'm happy (...) and because I like to be happy."
Retrieval	External Event / Past	Rely on a past event	"I have a friend who played with me at…"
Retrieval	External Event / Futur	Rely on a future event	"because Thursday I'm going to the park, I'm going to have a picnic in the zoo and there will be pancakes in the cantine"
Judgment/ Response	Missing scale description	The given answer is not in the scale	"between good and really good, I don't know how to choose…"
Judgment/ Response	Inconsistent choice	Choose a different answer in the scale than the spontaneous one	"not too often" *then she says* "almost never" *then she chooses* "no never"

Comprehension	No explanation	No explanation or justification given	
Judgment	Mood Expression / positive	Uses the smileys to express the mood	“there were those *(smileys)* who were happy, joyful, a little sad"
Judgment	Fun / positive	Finds the smiley funny	“Because the smileys are too funny"
Judgment	Too many options	Too many answer options to choose from	“there are so many good ones, I don't know what to choose”
Comprehension / Judgment / Response	Question	The question is easy to answer	“I found it easy to answer”
Retrieval	Familiarity	Know the smileys from somewhere	“I'm using the smileys a lot”
Response	Color	Mentions the color as beautiful	“I like the yellow color”
Judgment	Design / negativ	Doesn’t like the design	“because if there wouldn’t be the words to describe them, , I would not really understand what it meant.”
Judgment	Fun /negative	Finds the smileys not very funny	The heads weren't very funny (...) when they were very bad"

Chapter 4:

Designing Usability/User Experience Heuristics to Evaluate E-Assessments Administered to Children

1. Abstract

The application of electronic assessments (e-assessments) has risen recently, particularly among elementary-school-aged children. Despite their popularity, paper-based assessments are frequently converted into digital formats for efficiency gains, with little thought given to their user experience (UX) or usability. Individual differences, particularly among young children, can inhibit test-takers from completing the assessment tasks when the tasks are not designed to match their needs and abilities. Consequently, recent studies have raised concerns about the generalizability and fairness of e-assessments. Whereas heuristic evaluation is a standard method for evaluating and enhancing the efficacy of a product with respect to a set of guidelines, more information is needed about its added value when designing e-assessments for children. This paper synthesizes heuristics on the basis of the literature and expert judgments to accommodate children's abilities and ensure that all aspects of the domain are addressed appropriately. We present a final set of 10 heuristics, validated and refined by applying a heuristic evaluation workshop and collecting 24 expert surveys. The results indicate that the derived heuristics can help evaluate the UX and usability-related aspects of e-assessments with 6- to 12-year-old children. Moreover, the present paper proposes recommendations for a framework for developing usability heuristics that can be used to help researchers develop domain-specific heuristics in the future.

Keywords: User Experience, Usability, Heuristic Evaluation, E-Assessment, Children

2. Introduction

Over the last 2 decades, a dramatic shift has occurred in how young students are evaluated. The use of electronic assessment (e-assessment), also known as Computer-Based Assessment (CBA), Computer-Assisted Assessment (CAA), or Technology-Based Assessment (TBA), has become more prominent in the education sector. Whereas digital technology has gained popularity in the education sector, it is more than simply a "means to an end" for accomplishing the objective of finishing an exam. The OECD's Program for International Student Assessment (PISA) is an example of an e-assessment solution that measures the skills and knowledge of 15-year-old students with the goal of evaluating education systems throughout the world (*About PISA*, n.d.). However, paper-based assessments are often converted into digital formats without considering the test-taker's user experience (UX) or improved usability from the test-taker's perspective (Parshall et al., 2000). When using e-assessment tools, unoptimized design can prevent users from completing learning and assessment tasks (International Test Commission, 2006; Tselios et al., 2008). Additionally, individual differences in computer literacy (i.e., ability to handle technology) might impact learners' performance (Siddiq et al., 2017) and result in different outcomes that are not related to the problem-solving task (Lehnert, 2019; Weinerth, 2015).

Consequently, concerns have been raised about the generalizability and fairness of e-assessments (Weinerth, 2015). Therefore, e-assessment tools must be appropriately evaluated to understand the technology's potential impact on the test-taker. Evaluating new digital technologies in general, and e-assessments in particular, involves careful investigations of the needs and abilities of the target users, as they bring along unique characteristics that require special attention. When evaluating e-assessment interfaces for children, it is essential to consider the system's usability specifically with respect to this user group. Usability focuses on product design, particularly the user interface (UI) as the "channel for interactions" (Weinerth, 2015), p. 20). It focuses on how aspects of the product design (rather than a person's ability to use a system) improve task performance (Weinerth, 2015). Therefore, the concept of UX has been proposed to fill in the incomplete notions of usability. UX defines perceptions and reactions as "the users' emotions, beliefs, preferences, perceptions, comfort, behaviors, and accomplishments that occur before, during, and after use" (International Organization for Standardization (ISO), 2018), definition 3.2.3).

Addressing UX is important because children have unique cognitive and developmental abilities (Piaget, 1955), and any technology designed for children should be based on a good understanding of these abilities. Moreover, experts in the area of Human-Computer Interaction (HCI) or, more particularly, Child-Computer Interaction (CCI), have proposed several methods for evaluating usability (Lehnert et al., 2022), with the most prominent one being user testing, which involves the representative users performing tasks with the technology (Markopoulos & Bekker, 2003). Researchers capture data from these participants using a mix of observation, interviews, and surveys, documenting any problems or issues with the interface or system (Markopoulos et al., 2008) and allowing them to derive recommendations for redesigns. However, user testing has the drawback of being time-consuming and costly because it can require inviting several users (e.g., to a lab) to conduct the study (Lallemand & Gronier, G., 2018). In addition, testing with participants may require special preparation, especially with children, so a great deal of consideration and planning must be put into the testing procedure (e.g., simplifying instructions, providing additional comfort levels, and ensuring that the pace of the test can accommodate different attention spans) (Hanna et al., 1997).

Additionally, user studies in the context of e-assessment are challenging to implement, as any interference during an assessment can affect student performance. Moreover, simulating an e-assessment context is a complicated endeavor because the researcher has to control a combination of environmental factors (e.g., the presence of multiple students, the classroom or setting in general, and the feeling that one is being evaluated) to safeguard ecological validity. Therefore, alternative evaluation methods may be required when evaluating e-assessment applications that are designed for children.

The Usability Inspection Method (Nielsen & Mack, 2014) is often used as an alternative to user studies, especially when there are constraints on time or budget. This set of methods focuses on improving the design of an interactive system by having experts predict potential usability problems. There are various inspection methods, but the most widely cited and used is the heuristic evaluation (Nielsen, 1992, 1994; Nielsen & Molich, 1990). Similar approaches have been proposed, such as the cognitive walkthrough (Lewis & Wharton, 1997) and ergonomic criteria (Bastien & Scapin, 1995). However, heuristic evaluation can identify structural problems in the interface by looking for breaches of heuristic principles suggesting potential usability problems (Quiñones et al., 2018). The heuristics are rules or "guidelines" that evaluators use to conduct a heuristic evaluation. Evaluators inspect the interface, noting problems, and these problems are mapped onto heuristics, and a severity rating is attached to the problem on the basis of frequency and criticality (Nielsen & Molich, 1990). The method was initially proposed by Nielsen & Molich (1990), along with a set of general heuristics that are suitable for evaluating user interfaces (Nielsen, 1992, 1994). Nevertheless, domain-specific heuristics have been devised in cases when these generic heuristics are insufficient to cover all domain features of the particular system being evaluated.

Whereas heuristic evaluation is a comparatively quick and lower cost way to measure and improve a product's usability (Nielsen, 1992, 1994), tailoring the heuristic evaluation process to accommodate CCI requirements is necessary to ensure that all domain features are properly addressed. Research has confirmed that, to find a sufficiently large number of usability problems, researchers need to extend or modify Nielsen's generic heuristics to make them applicable to the features of the target interactive system (Markopoulos & Bekker, 2003; Quiñones et al., 2018) (e.g., social network (Quiñones et al., 2020)).

Looking into the trend of developing domain-specific sets of heuristics, some previous research has evaluated e-assessment platforms (Sim, 2009; Sim et al., 2007, 2008; Sim & Read, 2016). However, such work has not focused specifically on the child population or on the specific age-related features of the e-assessment application. The present paper addresses this gap. By developing domain-specific usability heuristics for children, we contribute by (a) reviewing the literature on the intersection of e-assessment, children, and usability/UX, (b) synthesizing a set of heuristics to conduct a heuristic evaluation on the basis of previous work and expert evaluations, and (c) proposing how to advance a framework for developing usability heuristics.

This article presents 10 heuristics for evaluating e-assessments in an educational context with children between the ages of 6 and 12 years. We propose these heuristics as a synthesis of related guidelines and heuristics from a literature review, following the methodology presented by Quiñones et al. (2018). We subsequently evaluated the heuristics with two rounds of validation: a heuristic evaluation workshop and an expert survey.

In the current paper, we define the building blocks that underlie the development of the heuristics and provide a review of the related literature. In the same section, we introduce different models for developing domain-specific usability heuristics and explain our choice in more detail. We present the methods and results from the chosen methodological approach for developing usability heuristics.

Finally, we discuss the results from the validation studies and future directions for designing usability heuristics for children's e-assessments. In the limitations and future work, we further propose improvements for the chosen methodological approach for developing the usability heuristics.

3. Related Work

Whereas various sets of usability heuristics exist for nonspecific populations, for the design of both general products and more specific use cases, to the best of our knowledge, no heuristics have been designed at the intersection of the following three components: the *children* (representing the particular user group), the *usability/UX,* and the *e-assessment* (the system component or domain context).

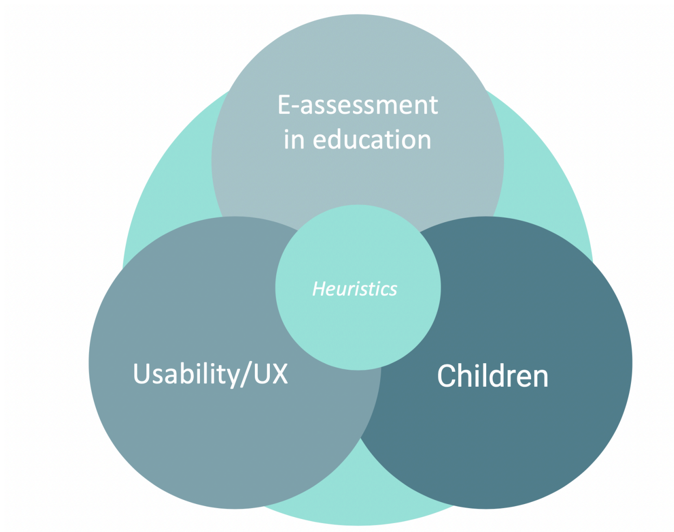


Figure 1. Three intersecting components for the development of the usability/UX heuristics.

In the following sections, we outline the three intersecting components and existing work in this area in more detail.

3.1. E-assessments in education

It is common practice for students to use computers to access course materials and complete coursework or assignments; however, they sometimes must take paper examinations, which interrupt their learning experience. Ideally, they should have a consistent learning experience, and there are benefits of delivering assessment content via computers or electronic delivery platforms. The benefits for students include the ability to study at their own pace, repeat incorrectly answered questions, and receive immediate feedback (Loewenberger & Bull, 2003), enabling them to identify areas where they need to develop and make improvements (Tsai & Oppenheimer, 2021). Interactive feedback approaches provided by technology-supported or electronic assessments can result in a positive learning experience, increased competency, and enhanced reflective practice among students (Wongvorachan et al., 2022). Consequently, the use of e-assessments can contribute to a more engaging and effective learning environment by altering how learning, teaching, and assessment occur, resulting in a more efficient and effective education system. There are numerous additional benefits for the various stakeholders (e.g., psychometricians, assessors, administrators, instructors, or tutors) (Sim, 2009).

Four general functions for e-assessments have been identified: formative, summative, certification, and evaluative (Hornby, 2003), from which the most common assessment categories can be derived: diagnostic, formative, and summative. A diagnostic assessment is conducted before or after an activity to ascertain the students' knowledge. Such assessments could be used to determine eligibility for employment or exemptions from a course of study. Formative assessment is carried out during the course to assess the efficacy of teaching and provide feedback to the students on their performance (Sim, 2009) (*assessment for students*). By contrast, summative assessment is administered to evaluate students' performances (Sim, 2009) (*assessment of students*). Students place the most emphasis on summative assessment because it determines their overall grade (Taras, 2002). This work focuses on evaluating summative e-assessments, as this type of assessment most closely resembles the implementation series and platform we used for validation. On the basis of the authors' work (Rowntree, 2015), we identified the following four purposes of assessment: selection, maintaining standards, and providing feedback to teachers and students.

Multiple-choice question (MCQ) tests and essays are still the most typical kinds of assessments utilized in public education (Graham, 2004). Nevertheless, 50 distinct assessment techniques are now in use in the field of education (Knight, 2001), such as multimedia (e.g., games and simulations for assessment purposes). Many techniques have inherent problems, whether administered via traditional methods or by computer; they strongly depend on the software platform on which they are delivered. From the literature, the following e-assessment elements were identified for a general student population (i.e., an age-independent population, which means we specify the age-related components at a later stage in the Method section):

1. **Independence:** E-assessments offer students the ability to study at their own pace (Loewenberger & Bull, 2003), or in the broader sense, to be able to communicate with the system independently;
2. **Immediate feedback:** E-assessments provide students with immediate feedback so that they can take correct actions (Loewenberger & Bull, 2003);
3. **Interactive elements:** Multimedia items enabled by technology can result in a positive learning experience, increased competency, and enhanced reflective practice among students

(Wongvorachan et al., 2022). We consider these elements necessary for a summative evaluation, as they can help students focus on a task.

The following features focus to a lesser extent on the *electronic* component but are still important considering the nature of summative assessments.

4. **Monitoring:** E-assessments enable students to identify areas in which they need to develop and to make improvements (Tsai & Oppenheimer, 2021) or, in our understanding, to identify issues with the interface and be able to learn from their mistakes;
5. **Selection:** E-assessments offer a basis for drawing judgments about students' performances or selecting candidates, for example, students who aim to enter a higher school need to be assessed to demonstrate that they are competent (Loewenberger & Bull, 2003);
6. **Maintaining standards:** To ensure a certain measure of the assessment (Rowntree, 2015), we focused primarily on the interface elements (not the content of the e-assessment);

We used these features as the basis for building our heuristics. We describe the further use of these features in the Method section.

3.2. Children as the target group

In the 1990s, as children became increasingly crucial as users and research participants for interaction designers and researchers (Druin, 1999b), the discipline of CCI began to grow consistently. CCI research focuses on the design of interactive technologies for children and the effects of technology on children and society (Hourcade, 2015). Designing for children is crucial because their cognitive, physical, and affective characteristics are very different from those of adults. For example, preschoolers have shorter attention spans and cognitive abilities that are not yet fully developed (Piaget, 1955). Therefore, to construct effective and engaging systems for a specific age group, designers must consider this group's developmental requirements, preferences, and abilities (Lehnert et al., 2022). Special design characteristics include simple and intuitive interfaces, visual aids and feedback, and plain instructions and guidance. Therefore, system designers must consider child-specific requirements, especially when valid conclusions must be derived from the children’s interactions with a system, such as an e-assessment platform. Only then will children be able to fully engage with the system and demonstrate their knowledge and skills accurately, thus contributing to the validity of the results from these assessments.

In evaluations of digital technologies with children, the children are often described as users (Druin, 1999a). Therefore, we need to build technology that is intuitive and user-friendly so that it can meet the demands of this specific user group. Designers may create technology that is more effective and accessible to all users by considering their capacity, needs, and prior experiences (e.g., computer literacy). It is especially critical in the setting of e-assessment because technology created with the user in mind promotes an experience that is less fraught with friction and errors, thus making it easier for users to concentrate on the tasks at hand (Avouris et al., 2001; International Test Commission, 2006). We focus on 6- to 12-year-olds because (a) this age group imposes requirements for the design of the heuristics that differ significantly from those of the adult context, as addressed by other authors (Hanna et al., 1997; Romeo et al., 2003), and (b) this age group is in urgent need of research for e-assessments (Al-Emran et al., 2018), particularly with regard to addressing concepts such as UX and usability in e-assessments (Ortiz-López et al., 2023).

We further decided to exclude children younger than 6 years as these user groups are still developing the necessary digital literacy skills. Consequently, assessments of these younger groups of children tend to differ significantly from assessments of school-aged children (as formal schooling begins at age 6 in most countries). In Piaget's stages, our study's target group is classified as being in the concrete operational stage (6 to 12 years) (Piaget, 1955). Children's numerical, categorical, and spatial abilities increase significantly in this age group. However, their mental manipulation is still tied to concrete things, which could affect how they perceive and interact with certain interface elements. Hence, despite the notable variations in features within the 6- to 12-year-old age range, there is a degree of homogeneity in specific characteristics, such as mental manipulation capacities. From the age of 12 and beyond, the distinction between designing for adolescents and adults becomes increasingly blurred. Adolescents frequently prefer items or activities that are intended for an adult population. As per Piaget's theory, they possess the cognitive ability to engage with abstract concepts like adults (Piaget, 1955).

3.3. UX and Usability

Two key concepts for establishing how successfully an interactive system addresses users' needs are usability and UX. There are many definitions of usability. One example refers to how easy it is to utilize a system and perform specific tasks while interacting with it through its user interface. According to the

ISO-9241-11, usability concerns the use of a product "by specified users to achieve specific goals with effectiveness, efficiency, and satisfaction in a specified context of use" (International Organization for Standardization (ISO), 2018) (definition 3.1). The idea of UX expands on the concept of usability by including a person's perceptions and responses as a result of the use or anticipated use of a product, system, or service (International Organization for Standardization (ISO), 2018) (definition 3.2.3). UX encompasses all of a user's emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviors, and accomplishments that occur before, during, and after they use the system. It is generally assumed to cover two dimensions: the pragmatic and the hedonic (Hassenzahl, 2001). Pragmatic UX (also known as ergonomic UX (Hassenzahl, 2001)) is concerned with instrumental or utilitarian elements and frequently prioritizes efficacy and efficiency in reaching a certain goal (Burmester et al., 2002).

On the other hand, the focus on hedonic components focuses on aesthetics (Diefenbach et al., 2014) or feelings, such as the joy of using a product (Hassenzahl & Tractinsky, 2006), which may be less relevant in the particular context of summative assessment than the pragmatic dimension. Hedonic components might be especially suitable for stimulating learning and motivating users to continue using a learning or assessment tool, which may benefit formative assessment. Hence, this paper concentrates on pragmatic UX and usability, with the goal of improving the system design to provide an appropriate experience for children while they perform a specific activity (Weinerth, 2015). To develop our heuristics, we considered the usability attributes proposed by the ISO 9241 standard (International Organization for Standardization (ISO), 2018): *effectiveness and efficiency.* We further addressed three additional usability concepts defined by Nielsen (Nielsen, 2012): *learnability* (i.e., how easy it is for users to accomplish basic tasks the first time they encounter the design), *memorability* (i.e., when users return to the design after a period of not using it, how easily they can reestablish proficiency), and *errors* (i.e., how many errors users make, how severe these errors are, and how easily can they recover from the errors). The last term is partly covered by the effectiveness definition by ISO; however, as the concept *of errors* can prevent mistakes and help the user recover from them and can therefore affect the validity of the results, we list it separately. We further considered one concept related to UX (Morville, 2004): *accessibility*. We chose to include accessibility as a UX concept, as e-assessment systems should offer fair chances to a variety of students (in mixed classrooms, which could consist of students with light disabilities, e.g., vision or hearing impairments). In the Method section, we outline all the concepts and their relevance in more detail.

3.4. Related heuristics and existing guidelines

A considerable amount of research relating to sets of heuristics and guidelines for the areas covered in Figure 1 has been published. Heuristics and guidelines have been used interchangeably over the years, and the majority are concerned solely with usability (i.e., Ben Shneiderman's Eight Golden Rules of Interface Design (Shneiderman, 1998)). More recent heuristics have been designed with multiple domains in mind; for instance, Susan Weinschenk and Dean Barker's usability guidelines (Weinschenk & T. Barker, 2000), including Nielsen's (1992, 1994), and they did a massive card sorting study that resulted in 20 of their heuristics. Several sets of heuristics have a UX focus (i.e., Arhippainen's UX heuristics (Arhippainen, 2013) and Colombo & Pasch's UX heuristics (Colombo & Pasch, 2012).

There are guidelines on the intersection of usability/UX and children (e.g., UI guidelines for children's computer products (Druin, 1999b) and Qualitätsmerkmale und Gestaltungsempfehlungen from the book Usability für Kids (Liebal & Exner, 2011)). Nielsen and Norman Group extracted 156 guidelines for UX design for children (age range: 3 to 12 years) on the basis of empirical work (e.g., usability testing). Across all their studies, children tested more than 80 websites and 36 apps, as well as mainstream websites. The studies took place in the United States, China, and Israel in participants' homes, schools, and usability labs (Liu et al., 2022). Another study created 350 guidelines grounded in a literature search on child development and children's technology, focusing on 5- to 8-year-old children (Gelderblom, 2008). This exhaustive set of guidelines summarizes knowledge from HCI, CCI, developmental psychology, and cognitive psychology. The guidelines were then classified into six categories: developmental appropriateness, specific skills (i.e., mathematics, built-in support, collaborative use, diversity of users, and interaction environments). Additional usability and UX guidelines were presented as 42 Design Principles for Children's Technology (Chiasson & Gutwin, 2005). However, they appeared to be missing relevant aspects compared with the guidelines from Gelderblom (2008). Domain-specific heuristics for children's e-learning applications (Alsumait & Al-Osaimi, 2009) were created; however, these heuristics did not cover aspects of e-assessment.

The final component, e-assessment in education, was almost exclusively composed of work by Sim (Sim, 2009; Sim et al., 2007, 2008; Sim & Read, 2016), except for one earlier manuscript (Squires & Preece, 1999). The heuristics here tended to be evidence-based, and the research studies were combined with a literature review to extract 11 CAA heuristics for usability evaluation. Sim (2009) validated the heuristics by conducting heuristic evaluation workshops with 32 HCI students. The results showed that the heuristics effectively identified usability issues in CAAs, even among novice evaluators.

It is evident that existing sets of heuristics identified in the literature need to focus more on young children, as these heuristics were primarily developed to assess adult user interfaces. Therefore, additional research is required to synthesize heuristics that are appropriate for evaluating the usability of e-assessments for young children. This research aims to ensure that e-assessments are effective and equally accessible to all users, thereby safeguarding the validity of the exam results. Sim (2009) argues that there are two main areas related to heuristics: the development of heuristics for specific domains and the analysis of methods to improve the efficiency of heuristics. The main objective of this research article is the development of heuristics. However, we also aim to provide insights and recommendations on how to use our methodology to retrieve the heuristics. Based on the gaps in the literature, our *study is aimed at developing an emerging set of usability/UX heuristics for evaluating summative e-assessment systems targeted at school-aged children (6 to 12 years of age).*

3.5. Current frameworks for developing heuristics

Research on developing heuristics has shown that many commonly used methods are extensions or adaptations of Nielsen's heuristics (1992, 1994). By contrast, other methods involve the creation of entirely new sets of heuristics tailored to a specific domain, such as e-assessment (Alsumait & Al-Osaimi, 2009; Quiñones et al., 2020). Whereas there are approaches to aid in the design of heuristics (Franklin, 2014; Hermawati & Lawson, 2015; Hevner et al., 2004; Hub & Čapková, 2010; Lechner et al., 2013; Rusu et al., 2011; Van Greunen et al., 2011), only a few authors have proposed a clear framework or protocol for developing and validating the heuristics (Jiménez et al., 2017; Quiñones et al., 2018). Hermawati & Lawson (2015) have critiqued and identified issues with the process of creating new sets of heuristics: (a) a lack of rigor, robustness, and standardization in the analysis of the effectiveness of domain heuristics and (b) significant deficiencies in efforts to validate new heuristics. Rusu (2011) offered an earlier framework for developing usability heuristics consisting of a six-step process. In a literature review, Quiñones & Rusu (2017) suggested that Rusu's (2011) methodology had an advantage when compared with other methods for developing heuristics, as it (a) presented clearly defined stages, (b) included a standard template for specifying heuristics, (c) included precise validation methods, and (d) could be applied iteratively. Moreover, Rusu's (2011) work expresses the importance of graphically showing each stage and including a validation stage.

Variations of the Rusu (2011) model targeting specific application domains have appeared over the years, for instance, for collaborative augmented reality remote systems (Franklin, 2014) or for a public administration portal (Hub & Čapková, 2010). These approaches offered more explicit descriptions of the activities the researcher had to perform; however, they needed to be more descriptive of the validation stage (Hermawati & Lawson, 2015; Van Greunen et al., 2011). Other researchers (Lechner et al., 2013) presented a methodology for developing usability heuristics with a validation procedure that incorporated only real users, not assessors (experts). However, the problem with this type of validation is that the ultimate users may need more information and the skills to assess whether a heuristic is accurately provided.

A framework that built on Rusu's (2011) approach is the QRR methodology (i.e., a methodology for developing usability/UX heuristics, named after the authors Quinones, Rusu, and Rusu) (Quiñones et al., 2018). With the QRR methodology, the authors stressed the importance of the iterative nature of the method and explained that (a) some stages might be optional, (b) some stages might overlap because it is necessary to implement two stages together, and (c) one stage might end earlier and one might go back to an earlier stage (Quiñones et al., 2018). In their article, the authors explained how to apply each stage to create a new set of usability/UX heuristics for a specific application domain (see Table 1).

Table 1. Steps of the QRR methodology to develop usability/UX heuristics (Quiñones et al., 2018), extracted from (Quiñones et al., 2020).

Step	Name	Description
1	Exploratory stage	Perform a literature review to collect information about the specific domain, their features, the usability/UX attributes, and the existing set of heuristics (or other relevant elements).
2	Experimental stage	Analyze data that were obtained in different experiments to collect additional information that was not identified in the previous stage.
3	Descriptive stage	Select and prioritize the most important topics of all information that was collected in the previous stages.
4	Correlational stage	Match the features of the specific application domain with the usability/UX attributes and existing heuristics (or other relevant elements).
5	Selection stage	Keep, adapt, or discard the existing sets of usability/UX heuristics that were selected in Step 3 (or other relevant elements).
6	Specification stage	Formally specify the new set of usability/UX heuristics using a standard template.
7	Validation stage	Validate the set of heuristics through several experiments (heuristic evaluations, expert judgments, user tests) in terms of their effectiveness and efficiency in evaluating the specific application.
8	Refinement stage	Refine and improve the new set of heuristics on the basis of the feedback that was obtained in Step 7.

Another very similar methodology to the QRR is the PROMETHEUS model (Jiménez et al., 2017). PROMETHEUS is also based on Rusu's (2011) work, however, with the difference that QRR (Quiñones et al., 2018) provides the expert opinion as a means of validating the freshly constructed heuristics.

4. Methodological Approach using QRR

The decision was made to use the QRR methodology (Quiñones et al., 2018) to develop our domain-specific heuristics because it provided the most detailed step-by-step procedure, including specific input and output strategies in each step. Additionally, it includes expert opinion as a validation tool. The QRR has also been widely applied and evaluated in recent research studies, for instance, an online course evaluation (Jahnke et al., 2021), accessibility of statistical charts for people with impaired vision and color vision deficiency (Alcaraz Martínez et al., 2022), social network UX heuristics (Quiñones et al., 2020).

An adaptation of the QRR methodology (see Figure 2) was used to develop the heuristics. In the first iteration, we walked through the complete steps of the QRR methodology. See Table 1 for a brief explanation and a complete description of the author's article (Quiñones et al., 2018). Each step and the results from it are presented below.

We describe our approach in line with the steps of the QRR methodology as described in Figure 2, with two iterations in the validation step. As the guidelines we based our heuristics on were predominantly evidence-based (Liu et al., 2022; Sim et al., 2008), we decided to merge Steps 1 and 2. As a validation methodology, we conducted a heuristic evaluation workshop. We then refined the set of heuristics (i.e., based on the outcomes from the heuristic evaluation). We validated the new set of heuristics with an expert judgment (i.e., a survey for the second round of validation). The outcomes from this last validation are presented as recommendations for the future development of usability heuristics for children in an e-assessment context. In the following sections, we describe each step illustrated in Figure 2, including the two validation rounds.

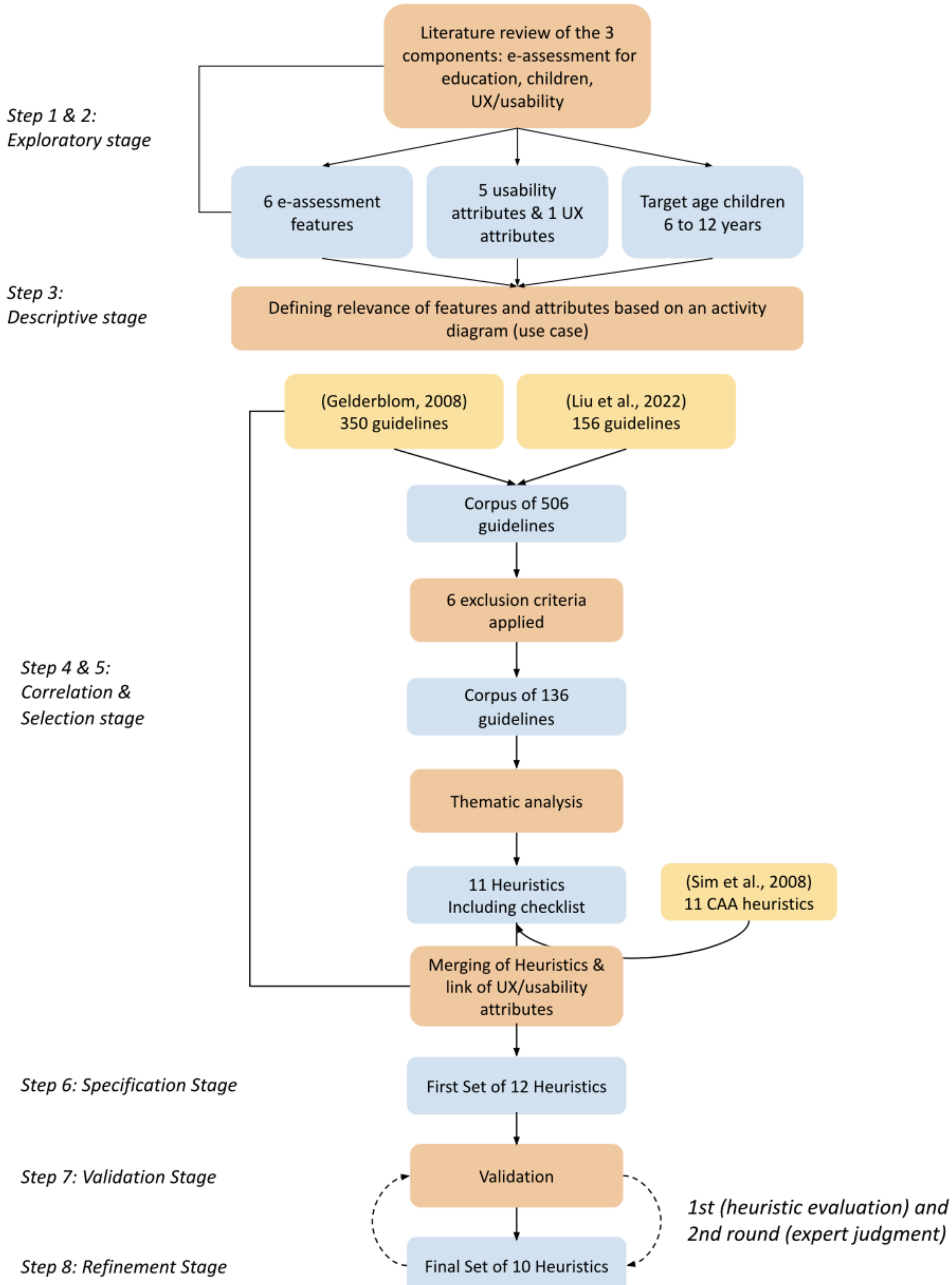


Figure 2. Illustration of the selection process used to create the final set of heuristics (input activity in orange, output activity in blue, referenced guidelines and heuristics in yellow).

Step 1: Exploratory stage

The first author conducted a literature review to collect features and attributes underlying the three components as illustrated in Figure 1: *e-assessment in education, usability/UX, and children*. The outcome from this investigation revealed a set of six e-assessment features (i.e., *independence, immediate feedback, interactive elements, monitoring, selection,* and *maintaining standards*) as described in Section 3.1. and five usability attributes (i.e., *effectiveness, efficiency, errors, memorability, learnability*) and one UX attribute (i.e., *accessibility*) as described in Section 3.3. We further narrowed our target group to children between the ages of 6 and 12, explained in detail in Section 3.2. In the next step, we searched the literature for existing heuristics, addressing the intersections of our three components from Figure 1. We found six relevant sets of heuristics with a focus on usability/UX (Arhippainen, 2013; Colombo & Pasch, 2012; Nielsen, 1992, 1994; Shneiderman, 1998), one heuristic with focus on the intersection of usability and CAA (Sim, 2009), and two on the intersection of usability/UX and e-learning (Alsumait & Al-Osaimi, 2009; Squires & Preece, 1999), four guidelines at the intersection of children and usability/UX (Chiasson & Gutwin, 2005; Druin, 1999b; Liebal & Exner, 2011; Liu et al., 2022), and one with a focus on children and technology (Gelderblom, 2008). In Section 3.4., we describe the above guidelines and heuristics in more detail. For the final selection of the guidelines that would form the basis of the development of the new heuristics, we were mainly concerned with the three intersections in Figure 1. Consequently, we decided to use the guidelines from Liu et al. (2022) as a first source because they addressed the intersections of children and usability/UX and because they were (a) evidence-based, (b) adapted/constructed from recent studies (both field and lab), and (c) internationally conducted in different cultures and populations. In addition, we selected Gelderblom's (2008) guidelines, as they were built on an exhaustive literature in the field of child development and children's technology (e.g., theories from cognitive and educational psychology). Finally, for the intersection of e-assessment and usability/UX, we selected the heuristics from Sim (Sim, 2009; Sim et al., 2007, 2008; Sim & Read, 2016). These were also evidence-based and based on a literature review in the field of CAA. As outlined earlier, the missing intersection of children and e-assessments was the core motivation for formulating a new holistic set of heuristics covering all intersections as illustrated in Figure 1.

Step 2: Experimental stage

Step 2, the experimental stage, was merged with Step 1, as the guidelines that we selected were evidence-based. Therefore, we decided not to conduct additional experiments.

Step 3: Descriptive stage

With Step 3, the descriptive stage, we synthesized the information we collected earlier in the literature review (i.e., defining underlying features of an e-assessment platform in more detail). For this purpose, we created an activity diagram to illustrate the users' interactions with a representative system. Building on an existing collaboration, we had access to an e-assessment platform developed by the Directorate of Evaluation, Foresight, and Performance (in French: Direction de l'évaluation, de la prospective, et de la performance; DEPP). DEPP assigned its primary activity to evaluating and measuring skills in education and training with a yearly large-scale Mathematics and French e-assessment. These assessments measure whether the students have achieved their educational objectives and are carried out through an application on a tablet. It contributes to assessing the policies conducted by the Ministry of National Education in France. The construction of the activity diagram for the e-assessment platform was aided by Sim's (Sim, 2009), which defined the most important elements.

- *ellipses* represent *actions*;
- *diamonds* represent *decisions*;
- *bars* represent the start (*split*) or end (*join*) of concurrent activities;
- a *black circle* represents the start (*initial node*) and end (*final node*) of the workflow;
- arrows run from the start toward the end and represent the order in which activities happen;

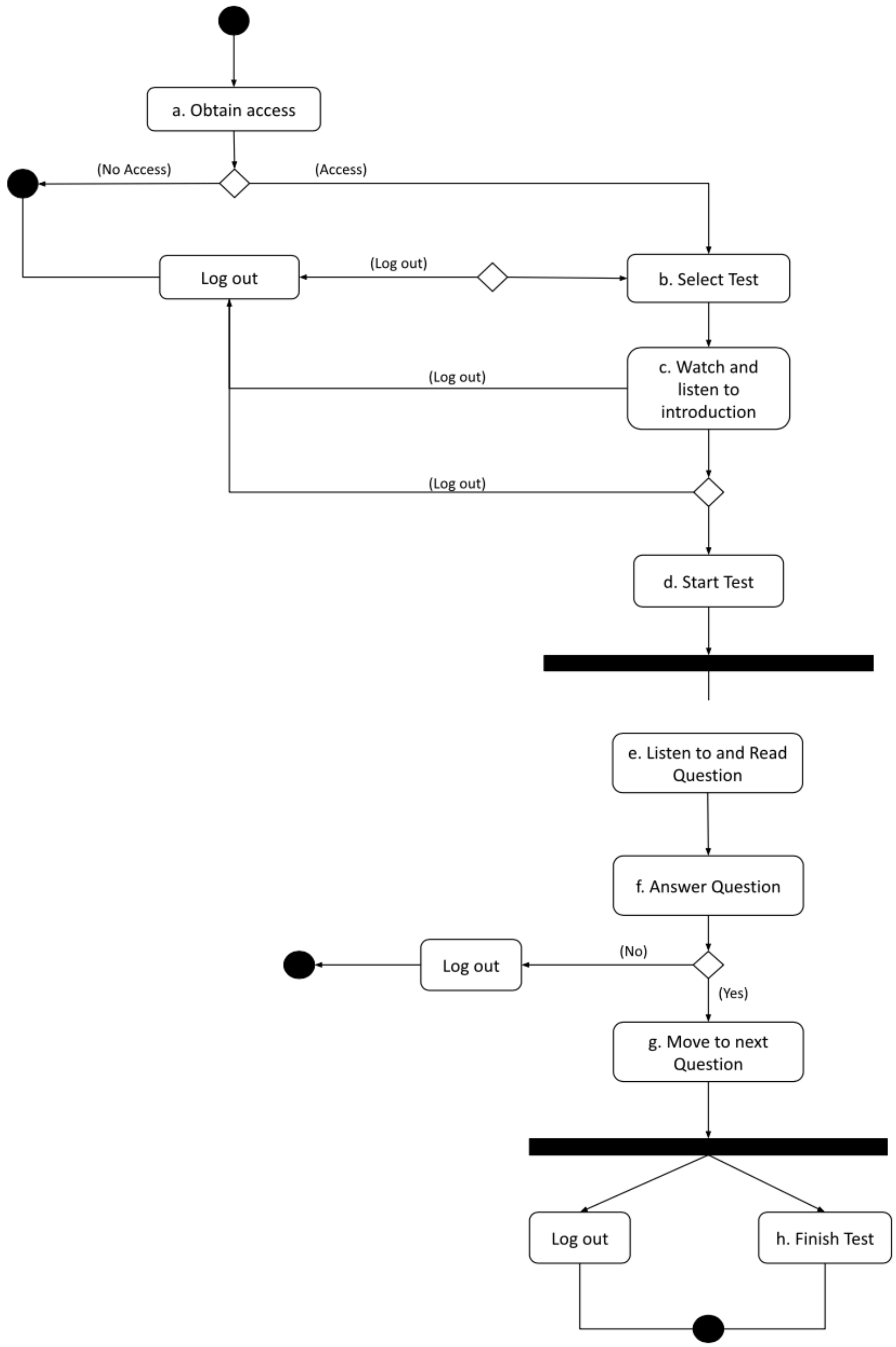


Figure 3. Activity diagram of how the user interacts with the DEPP application.

Step 4: Correlation Stage

Based on the activity diagram, eight e-assessment features (see Figure 3: a to h) were identified. In the next step, we wanted to verify whether these features were all represented in the previously defined e-assessment systems (see Section 3.1.) and usability/UX attributes (Section 3.3.). Therefore, as shown in Table 2, we matched the related features identified in Figure 3 to the attributes defined in Sections 3.1. and 3.3. This matching process was done by the first author and validated by another researcher to ensure appropriate categorization. An attribute was categorized as *very important* if it covered four or more features and as *somewhat important* if it covered fewer than four features. Two e-assessment features defined in the literature (i.e., selection and monitoring) were not covered by the DEPP application (Figure 3). The DEPP platform is summative in nature, and the students do not receive their test outcome, which means they were not able to identify areas that needed improvement or development (monitoring), nor were they (from the user's perspective) able to make judgments about their performance (selection). The two e-assessment features that were not covered are more relevant to other stakeholder groups (e.g., teachers). Overall, as can be seen from Table 2, all other features identified from the DEPP application were addressed.

Table 2. Relevance for e-assessment, usability, and UX attributes based on the identified features from the activity diagram in Figure 3.

Topic	Relevance		
	Very important	**Somewhat important**	**Not covered**
E-assessment attributes	*Independence* (Loewenberger & Bull, 2003) (a, b, c, d, g, h); *Immediate feedback* (Loewenberger & Bull, 2003) (a, b, e, g, h); *Maintaining standards* (Rowntree, 2015) (a, b, e, f, g, h);	*Interactive elements* (Wongvorachan et al., 2022) (c, e, f);	*Selection* (Loewenberger & Bull, 2003)*, Monitoring* (Paige Tsai & Danny Oppenheimer, 2021)
Usability attribute	*Effectiveness* (International Organization for Standardization (ISO), 2018)(a, b, c, d, e, f, g, h); *Efficiency* (Nielsen, 2012) (a, c, e, f); *Errors* (Nielsen, 2012) (a, b, c, d, f, g, h);	*Memorability* (Nielsen, 2012) (c, f); *Learnability* (Nielsen, 2012) (c, f);	
UX attribute	*Accessibility* (Morville, 2004) (a, b, c, d, e, f, g)		

An example from the DEPP application is shown in Figure 4. In this example, Steps e (i.e., listen to and read question) and f (i.e., answer question) from the activity diagram (Figure 4) are portrayed. Both of these steps represent the e-assessment attribute *interactive elements*, as the children are required to drag and drop the correct answer into a predefined field (i.e., Groups 1 or 2 in Figure 4). Moreover, this example represents the usability attribute *effectiveness*, as it is important that the student succeeds in correctly dragging and dropping the required response to the field. Finally, it is also related to the UX attribute *accessibility*, as the color contrast is not optimal (concerning students with visual disability problems, e.g., color blindness). Additionally, there is no way to (repeatedly) use an auditory question and response input, for example, to help users with reading disabilities (dyslexia) understand the question.

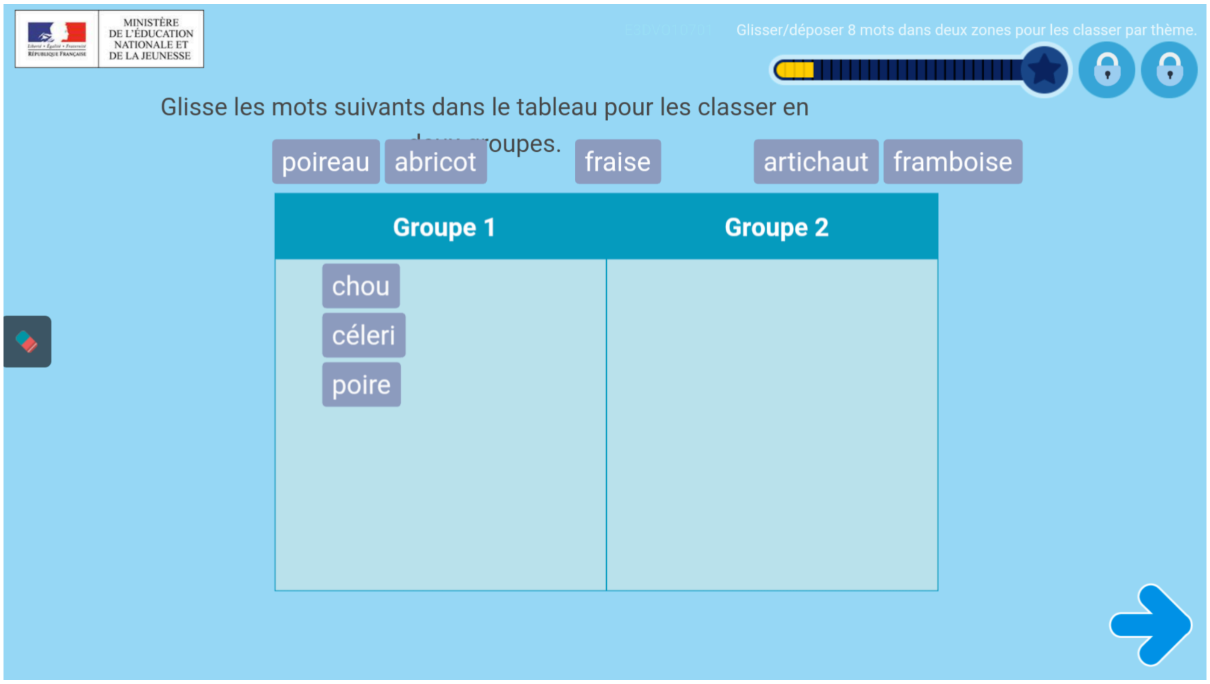


Figure 4. The DEPP application, representing Steps e and f from the activity diagram.

Step 5: Selection Stage

Based on the previous steps, an initial corpus of 506 guidelines (Gelderblom, 2008; Liu et al., 2022) and 11 CAA heuristics (Sim, 2009; Sim et al., 2007, 2008; Sim & Read, 2016) (see also Figure 2) were identified. The guidelines and heuristic contain different levels of details, which were then individually analyzed to make them comparable to the CAA heuristics in a later stage. For the 506 guidelines, we began our analysis by applying a set of exclusion criteria to remove guidelines that were not applicable. We excluded guidelines that:

- mainly targeted children **outside the age range of 6-12 years** (e.g., children younger than 6). For example, we excluded the following guidelines: "*For preoperational children (around age five), direct manipulation with the mouse is preferable to indirect manipulation through keyboard commands, as they lack the working memory capacity the latter requires*"(F.1.45 (Gelderblom, 2008)).
- focused on **devices other than tablets or computers** as an interaction technology, as these devices are currently the most common technologies for running e-assessment applications on. Therefore, for example, we excluded guidelines addressing Human-Robotic Interaction, as these forms of interaction differ greatly from traditional instruments. An example of such a guideline

is: *“To enhance personalised interaction between humans and robotic toys: The robot should be able to recognise the human user as an individual and it should be able to alter its behaviour according to characteristics and behaviour of that particular user etc.* (Gelderblom, 2008)*”*

- addressed serious developmental disabilities or handicaps in (Gelderblom, 2008) (e.g., blind children), as the designing of such specific interfaces had to follow different and more specific requirements, and we aimed to consider **basic design knowledge for all users** (i.e., 6 to 12 years). However, we did include design principles for children with (a) reading disabilities (e.g., dyslexia), (b) visual disabilities (e.g., color blindness), (c) cognitive disabilities, (d) motor disabilities, or (e) hearing disabilities, as designing for these users follows principles of accessibility in an educational context. An extract of an example for an excluded guideline is: *“...Blind children can use a joystick to move sounds to fixed positions in space, to ‘catch’ moving sounds and to ‘throw’ sounds...”* (Gelderblom, 2008)*;*
- addressed any specific differences between children regarding their **social background or gender** (e.g., culture, ethnicity, gender, demographics, social and economic situation). As an example, we excluded the following guideline: *“Make activities gender-neutral or gender adaptable: Girls prefer pretend play based on reality while boys prefer pretend play based on fantasy. Girls lean more toward education and strategy games and boys toward combat and sport games, etc.”* (Gelderblom, 2008)*;*
- cannot be linked to an **individual e-assessment context** (i.e., on a policy level) in a classroom context. For example, we excluded guidelines that addressed collaborative aspects, for example, *“Requiring children to move physical props around slows down interaction. Using more (and different) props together promotes collaboration as it avoids turn-taking and it encourages different collaboration styles.”* (Gelderblom, 2008)*;*
- did not concern the **interaction design field**, such as guidelines that referred solely to the content creation of an assessment item or the physical experience, for example: *“Present the breadth of the website's content on the homepage, to achieve a true initial impression of what the site has to offer.”* (Liu et al., 2022)*;*

To attain *intersubjective consensus* with respect to the exclusions, three authors performed the analysis. When in doubt about whether to exclude or include a guideline, they discussed any issues until they reached a consensus. This approach was selected to enhance the reliability of the findings (Miles et al., 1994).

After applying the exclusion criteria, the corpus that was left consisted of 136 guidelines. In the next step, the first author identified common themes in the existing guidelines (Druin, 1999a; Lavery et al., 1997) using a thematic analysis approach (Guest et al., 2011). The coding scheme was then reviewed by the other authors to attain *intersubjective consensus and* ensure the comprehensibility and use of the designed codes (Miles et al., 1994). The identified themes described the guidelines in one or two words. In the following step, we merged similar themes and framed the heuristics. Then we formulated the corresponding checklist points underlying every heuristic which included a detailed description of every heuristic checking for their applicability (Druin, 1999a; Lavery et al., 1997). The outcome of this process was an initial set of 11 heuristics with their underlying checklist. The first and last authors then compared the derived heuristics with the existing CAA heuristics (Sim et al., 2008) to identify missing elements. The CAA heuristics contained another level of detail, and it was for this reason that we did not match the heuristics earlier. Furthermore, we changed several formulations of the heuristics and added one more heuristic that referred to the CAA Heuristic 11 from (Sim et al., 2008) *"Minimise external factors which could affect the user".* The outcome of this stage was a first complete set of 12 heuristics, including their checklist items.

Step 6: Specification stage

In Step 6, the specification stage, and similar to the template for the QRR methodology (Quiñones et al., 2018), we added a detailed description to every heuristic along with a visual representation to convey the main message for every heuristic. Moreover, we formulated the benefits of applying the specific heuristic (Quiñones et al., 2018). Finally, we linked the e-assessment and usability/UX attributes to the heuristics (as defined in Table 2) to check for their completeness. We propose the first set of 12 heuristics, including (from top to bottom) the *heuristic's number, name, description, benefit, example, wireframe, checklist items, features, usability/UX attribute, source* (see table 3 for an overview and Appendix A for an detailed overview of the final heuristics).

Step 7a: First validation stage

Procedure for the heuristic evaluation

We conducted a heuristic evaluation workshop with two HCI experts with primary knowledge about designing e-assessment interfaces (both experts consented to take part in the study). One of the experts (second author) had additional experience with the methodology used for heuristic evaluation and design. We did not conduct more workshops, as it was challenging to recruit experts who were familiar with both usability and e-assessment. The aim of this workshop was to obtain an initial critique of the set of heuristics to refine the heuristics before conducting the main validation study. For the workshop, we used the same application that was used to construct the activity diagram in Figure 4, from the French Ministry of Education, DEPP, which aims to monitor the competence of 6 to 8-year-old students in French and mathematics. We used this application because it meets all the requirements for an application (i.e., summative assessment of 6 to 8-year-old children). The workshop followed Nielsen's (Nielsen & Molich, 1990) procedure, which entails an individual and collective evaluation. For the personal evaluation, each evaluator (a) read the heuristics and became familiar with them; then (b) examined and used the e-assessment application; then (c) wrote down any problems they encountered with the application; and then (d) linked the problems to the heuristics and rated their severity (i.e., from 1 = *low* to 3 = *severe*; (see the Appendix B for the individual form for the evaluators). For the collective part, the evaluators came together, and each evaluator (a) took their list of problems and (b) discussed it with the moderator (first author) who took notes on the common problems (see see Appendix C). This process resulted in an aggregated list of problems associated with the heuristics that had been violated, the frequency score (how often we observed the same problem in the application), and their severity scores with the following classifications (see Appendix C for the reporting form and outcomes):

1. **Cosmetic Problem:** The identified issue is mainly *aesthetic*. Correcting it will improve the interface's layout and make the tool more pleasant to use. It is not expected to affect the user's performance in the e-assessment, but it would be great to fix if extra time is available.
2. **Minor problem:** The severity of the issue is *medium,* and it can be overcome by the user with some extra effort. Correcting it will improve the usability of the tool and the overall experience. It might impact the test performance of some students (e.g., students with less computer/tablet literacy skills).
3. **Major Problem:** The severity of the issue is *important*, and it is expected to create difficulties for all users regardless of their literacy. Correcting it is expected to have a direct impact on the usability of the tool and the validity of the test; it could affect the results of the test, and fixing it should be given high priority.

Step 8a: Refinement stage

Results of the heuristic evaluation

In Stage 8a, the refinement stage, the heuristics were refined on the basis of the results and feedback from the heuristic evaluation workshop described in Step 7a. We briefly summarize the findings that came from the experts' reviews so that readers can understand why we changed certain heuristics. The experts reported a total of 33 incidents with the e-assessment platform, and all issues could be linked to a heuristic except for Number 8. Heuristic 6 was linked most often to an issue, followed by 7, 4, 3, and 2. The other heuristics were linked only once or twice to an issue but were reported to have high severity. For a complete overview of issues, frequencies, and severity, we refer readers to the Appendix C. As a result of this workshop and the underlying discussions with the experts, the first author refined and merged some of the heuristics, resulting in a total of 10 heuristics. First, we merged Heuristic 7, "*Visually highlight interaction design elements*," with Heuristic 2 by adding the *visual feedback* component. We defined the new heuristic as "*Ensure timely and visual feedback on children's actions.*" This was because the experts reported some overlapping characteristics between the two heuristics. Second, we removed Heuristic 8 "*Adjust interaction design elements to children's abilities,*" as the researcher did not see any added value of having this as a separate heuristic. However, we kept part of the information from the checklist items and moved it to the new Heuristic 7, *"Design age-appropriate interface elements."* Third, we added some of the checklist description of the old Heuristic 8 to the new Heuristic 8 and reframed it to *"Make the user interface intuitive and accessible to all children."* Finally, we added *external factors* to the last heuristic to put more emphasis on factors that influence the e-assessment experience from the outside, such as an unstable internet connection. For the second validation round, the expert survey, we repeated Stages 7 and 8.

Table 3. Overview of all e-assessment heuristics before and after the first validation round.

#	*Names from first set of heuristics (before first validation)*	#	*Names from second set of heuristics (after first validation)*	*Description of second set of heuristics*
1	**Prevent errors and help to recognize them**	1	**Prevent errors and help to recover from them**	Prevent errors with the interface from affecting test performance. If errors occur, help children to recover from their mistakes.
2	**Ensure adequate and timely feedback on children's actions**	2	**Ensure timely and visual feedback on children's actions**	Provide adequate and immediate visual feedback on children's actions and show them the effect it had on the screen.
3	**Provide support and guidance throughout the interface**	3	**Provide support and guidance throughout the interface**	Children profit from voluntary support. Scaffolding options should help the child navigate throughout the interface.
4	**Make children's actions and system status transparent**	4	**Make children's actions and system status transparent**	Children should get insights into the navigation process and system status.
5	**Allow the children to control the interface**	5	**Give children freedom to control the interface**	Give children the most flexibility possible for controlling the interface.
6	**Anticipate consistency and standards**	6	**Anticipate consistency across the interface**	We should strive for consistent interface design within and across the platform.
7	**Visually highlight interaction design elements**	7	**Design age-appropriate interface elements**	Make sure the interface matches standards in designing for children and addresses age-appropriate motor coordination for that age group.
8	**Adjust interaction design elements to children's abilities**	8	**Make the user interface intuitive and accessible to all children**	Make sure the interface is accessible to all children and that they can adjust it to their needs.
9	**Design age-appropriate interface elements**	9	**Account for diverse technology literacy**	Children might not have the same technology literacy, so we have to prevent them from making errors based on fewer experiences with the assessment interface or a particular device such as the tablet.
10	**Make the user interface accessible**	10	**Eliminate disruptive and external factors of the delivery platform**	The delivery platform should be responsive, secure, and robust.
11	**Account for diverse technology literacy**			
12	**Eliminate disruptive factors of the delivery platform**			

Step 7b: 2nd Validation stage

Expert Judgment Procedure

For the second iteration, we used LimeSurvey to design a survey that asked experts for their opinions about our heuristics (expert judgment). All experts filling in the survey had to consent to participating in the study. We created a questionnaire that assesses evaluators' perceptions of the new set of usability/UX heuristics (as presented in Table 3, second set) concerning the following dimensions as outlined in the QRR methodology (Quiñones et al., 2018):

- *D1 – Utility: How useful the usability/UX heuristic is;*
- *D2 – Clarity: How clear the usability/UX heuristic is;*
- *D3 – Ease of use: How easy it was to associate identified problems with the usability/UX heuristic;*
- *D4* – Necessity of an additional checklist: How necessary it was to complement the usability/UX heuristic with a checklist;
- Q1 – Easiness: How easy was it to perform the heuristic evaluation based on this usability/UX heuristics set?
- *Q2 – Intentions: Would you use the same usability/UX heuristics when evaluating similar software products in the future?*
- *Q3 – Completeness: Do you think the usability/UX heuristics cover all aspects of usability for this software product?*

After an initial pilot round (*N* = 4), we took out dimensions D4 and Q1. In the proposed model and survey approach by the QRR methodology (Quiñones et al., 2018), the experts were asked to apply a heuristic evaluation (i.e., in our case, to a particular e-assessment application). Due to recruitment challenges and time commitments, the survey was redesigned so that the experts were not required to evaluate an e-assessment application at this point, as they preferred to concentrate on reviewing the heuristics on the basis of their expertise. The details and recommendations for designing an expert judgment survey (i.e., to validate a new set of heuristics) is discussed in this paper's limitations and future perspectives.

The final survey was organized into five main sections:

1. **Introduction:** *the goal of the study, data collection, confidentiality,* and *consent declaration*;
2. **Participants' expertise**: current *job title*, degree of *experience* (very inexperienced to very experienced, 5-point Likert scale on HCI, CCI, CBA, usability/UX, education & learning), *has conducted* versus has not *conducted a heuristic evaluation* before, experience in *designing or validating a new set of heuristics*;
3. **The review of the heuristics**: an overview of all heuristics (name only) and first impression of *clarity* (5-point Likert scale ranging from 1 = *not clear at all* to 5 = *very clear*), ratings on four dimensions (*utility, clarity, simplicity,* and *relevance* with a semantic differential scale), and an open question asking for recommendations to improve the underlying heuristic;
4. **Overall ratings** of the heuristics: *completeness* (5-point Likert scale ranging from 1 = *strongly disagree* to 5 = *strongly agree*), *intentions to use* (5-point Likert scale ranging from 1 = *extremely unlikely* to 5 = *extremely likely*), other *application areas.*
5. **Final note:** Offering an incentive for study participation (i.e., a gift card for 35 Euros).

Step 8b: Refinement stage

Results of Expert judgments (survey)

A total of 24 participants were recruited via Linkedin and the snowball method, beginning with University and partnership contacts (e.g., e-assessment communities). The participants received a gift card worth 35 Euros as compensation. All participants further defined their job in one of the predefined categories, resulting in one designer, 14 researchers, two developer/IT specialists, three coordinators/managers, three education specialists, and one other, specified as a clinical psychologist. We further asked for participants' expertise in (a) designing heuristics and conducting a heuristic evaluation method, (b) research fields considered essential elements for our designed heuristics: Education & Learning (E&L), CBA, usability/UX, CCI, HCI.

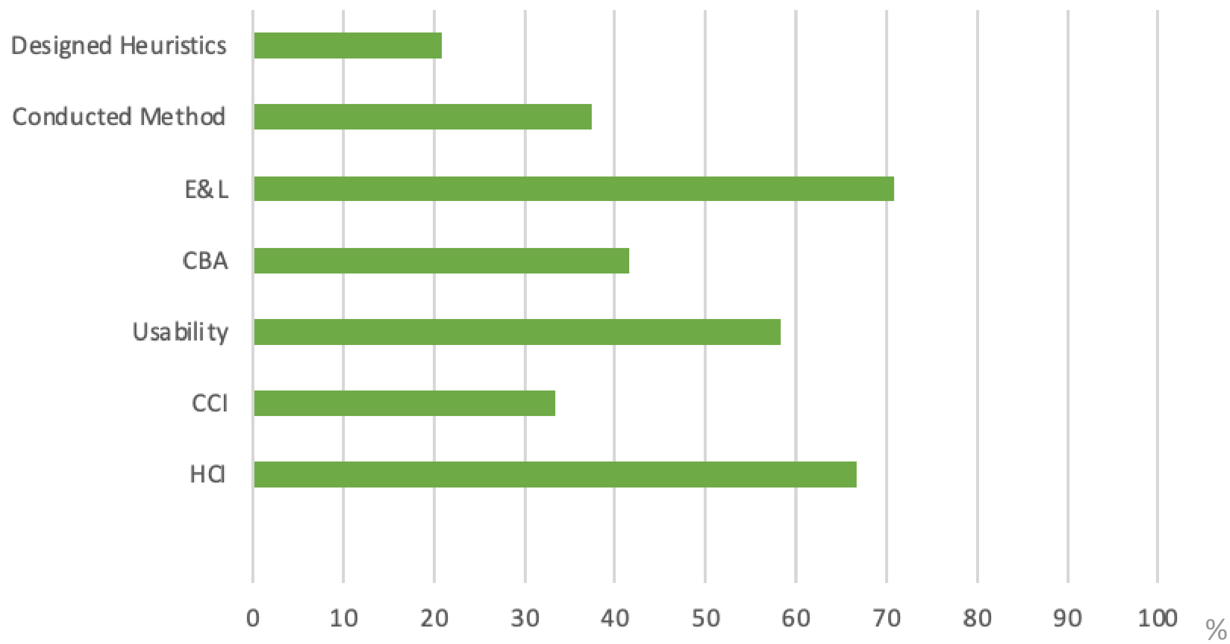


Figure 5. Overview of participants' expertise in percentages. The participants could choose multiple areas of expertise.

As shown in Figure 5, only 20.83% of participants had experience designing heuristics, and 37.5% had previously evaluated heuristics. Moreover, Figure 5 shows that the primary expertise was in the fields of E&L, HCI, and usability (more than 50% of the participants stated that they were either experienced or very experienced in one of the underlying fields). Less than half of the experts claimed to have expertise in CBA and CCI.

Before entering the specific evaluation of each heuristic, the participants were presented with an overview of all heuristics (without further descriptions or checklists). As a first impression, 20.8% rated the heuristics as somewhat clear, 62.5% as clear, and 16.7% as very clear.

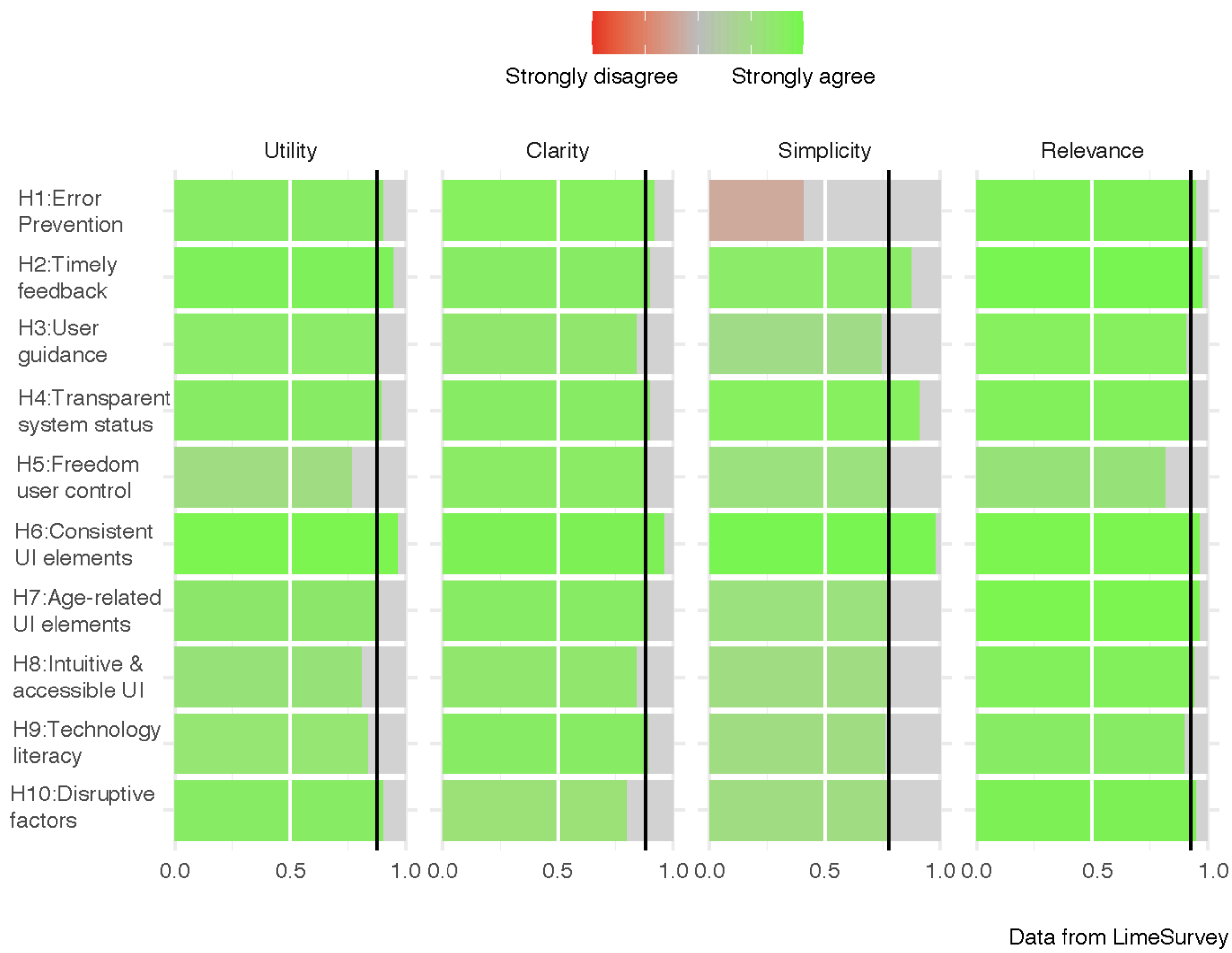


Figure 6. Overview of the mean ratings across all heuristics, including the average mean score per dimension (black lines).

Overall, the heuristics were rated as strongly positive, with mean scores as follows: utility = *4.37*, clarity = *4.41*, simplicity = *3.88*, and relevance = *4.65* (rated on Likert scales ranging from 1 = *strongly disagree* to 5 = *strongly agree*). We observed the lowest overall rating on the simplicity scale, with an outlier on Heuristic 1 (M = 2.01). We investigated whether having previous experience made a difference in the extreme rating on the simplicity dimension of Heuristic 1. However, conducting versus not conducting a heuristic evaluation did not explain any difference in the Heuristic 1 simplicity ratings (U = 40.500, Z = -0.393, p = .695). Regarding experience with designing heuristics, we could not say whether there was any relationship with the rating of simplicity for Heuristic 1 because very few people had experience designing heuristics (n = 5 vs. n = 16 who had no experience). The results of a Mann-Whitney U test showed no effect (U = 20.000, Z = -1.756, p = .079).

Additionally, there was no effect of relevant expertise (i.e., in the fields of HCI, CBA, etc.) on the simplicity rating for Heuristic 1. Therefore, an order effect could have caused the low ratings on this dimension on Heuristic 1. Participants were still getting used to the structure of the heuristics as presented in the survey, and they might have experienced it as complex at the beginning. Additionally, The extreme rating could also have been caused by a misunderstanding of the wording of that dimension: complex versus simple (i.e., Is the text that describes the heuristics experienced as difficult or straightforward, or is the process of applying Heuristic 1 expected to be complicated or simple?).

5. Discussion

In the following section, we discuss the outcomes and future perspectives for the heuristics we designed, and we provide recommendations for improving the existing methodology to develop them.

5.1. Generalizability versus specificity of heuristics

When designing heuristics, researchers often question their applicability and generalizability. As initially identified in the literature, a heuristic is a common rule of thumb that is supposed to help designers decide and critique existing interface designs. They were not designed with a specific target group or domain in mind and should be "broadly applicable" (Nielsen & Molich, 1990). However, recent research has criticized this proposed generalizability by claiming that not all usability errors can be detected with broadly applicable heuristics, as critical aspects are overlooked (Markopoulos & Bekker, 2003; Quiñones et al., 2018).

In line with this shortcoming, one critique we received during our evaluations was that several heuristics could apply to a broader audience. For example, Heuristic 1 (i.e., error prevention) could account for participants with lower computer literacy skills in general (e.g., the elderly population). Hence, the first five heuristics we designed were closely related to the 10 general principles for interaction design (Nielsen, 1992, 1994). Therefore, these were incorporated into the set of heuristics, as there must be a common denominator for all interface designs, regardless of the target group and domain. This need for a common denominator is also why we decided to specify the heuristics in more detail by including a checklist. The checklist provides the researcher with more explicit points for evaluating the usability of children's e-assessments. For example, in Heuristic 1, error prevention, previous research (Liu et al., 2022) has reported that children often tend to quit applications by mistake. Error prevention has been discussed in broader terms in educational psychology, for example, during emergent literacy learning (Gabas et al., 2023). At this juncture, three strategies have been proposed to respond to errors: (a) accept the logically correct response, (b) correct the error and discuss it, and (c) lead the child to the correct response (Gabas et al., 2023). For example, designers could provide a pop-up window in the e-assessment application asking children to confirm that they want to leave before it logs them out of the system. By doing so, children can understand and correct their actions and learn from their mistakes for future interactions with the application. Therefore, researchers have pointed out the importance of implementing error-handling strategies to scaffold children's knowledge and understanding (Henderson et al., 2002), thus reducing future errors in the process. The child-specific error handling strategy is an example of an extension of the general heuristic by Nielsen. It addresses the specific needs and requirements for interface design for young children. However, more research is needed to investigate whether these proposed strategies could account for other target groups (e.g., the elderly population) and whether similar implementation strategies (i.e., checklist items) can be formulated.

5.2. Addressing different technologies and hardware devices

As for the technological systems used for e-assessment, integrated learning management systems (LMS; e.g., Moodle) of the institutions are used the most, followed by bespoke software (i.e., all those accessible through the network, e.g., Kahoot). The last category forms the mobile technologies (tools or applications for exclusive use through mobile devices) with applications focused on assessment, for instance, ExamSoft (Ortiz-López et al., 2023).

Additionally, recent trends involve discussions of the possibilities for Robotic, Augmented Reality (AR), or Virtual Reality (VR) interventions (Gelderblom, 2008). We focused solely on tablets and computers as possible interaction devices in the current set of heuristics. We excluded literature that addressed, for example, VR or AR because the interaction modes are incredibly different from traditional instruments, and therefore, many of the heuristics would not be applicable. However, other research has been conducted in this section; for instance, (Endsley et al., 2017) developing a set of heuristics to advance AR design in multidimensional augmented environments. Additionally, we decided to exclude VR or AR technologies from the current heuristic design, as they face several limitations (i.e., financial feasibility (Riley & Stacy, 2008); poor instruction design (Chen et al., 2005); physical and psychological discomfort

(Al-Azawei et al., 2019). Future research has to focus on developing e-assessment heuristics for this particular technology.

Moreover, we decided to address only some differences between tablet and computer devices in our heuristics. This decision came primarily from the fact that we wanted the heuristics to be applicable to both devices, as the sheer literacy differences across children can influence their experience. Consequently, researchers found that more exposure to a particular device (i.e., a tablet or computer) has been shown to significantly impact a user's preference for and efficiency with a specific device (Gelderblom, 2008). Research indicates that although touchscreens are an excellent input device for younger children because of the direct correlation between the child's actions and the onscreen effect, touchscreens are not always more efficient than mouse control (Scaife & Bond, 1991). Young children (ages 3 to 7 years) in particular seem to have difficulty selecting, dragging, and moving objects around the screen using a finger. Therefore, children in this age group have generally fared better when using a mouse (Romeo et al., 2003). A study has confirmed that the touch screen seems advantageous, but once children become proficient with a mouse, they do better with a mouse (Ager & Kendall, 2003). Therefore, we should investigate this relationship in more detail in future studies, as children nowadays have more exposure to touchscreen devices (i.e., smartphones).

Future research on the development of e-assessment heuristics for children should address how lower levels of literacy with a certain device or type of technology influences the e-assessment experience.

5.3. Recommendations for the methodological approach

The QRR framework (Quiñones et al., 2018) has been designed in extensive detail and comes with a significant number of examples from different areas of application, for instance, social networking (Quiñones et al., 2020). We recommend that authors provide an overview of these sources and studies to help other researchers understand and look into the particular example reports (i.e., use cases) that are closest to their topic of interest. The inclusion of example reports has the potential to enhance the accessibility and practicality of the complex approach, which may otherwise be perceived as daunting and may be discouraging for researchers. Second, in the first step, the exploratory stage, we recommend that authors include an analysis of the target audience in addition to the focus on the domain. As discussed in the previous section, the target audience is particularly important in fields where several different target groups are found with unique characteristics that are worth addressing. In addition, the technology of the target system should be considered, as it also influences the selection and definition of the evaluation criteria. Figure 7 presents a summary of the main building blocks, which should form the initial steps of the exploratory stage. The *usability/UX attributes* and *topic/domain of interest* are the building blocks that are currently included in the QRR methodology; consequently, we recommend that the type of *technology* and *user groups* be added as outlined below.

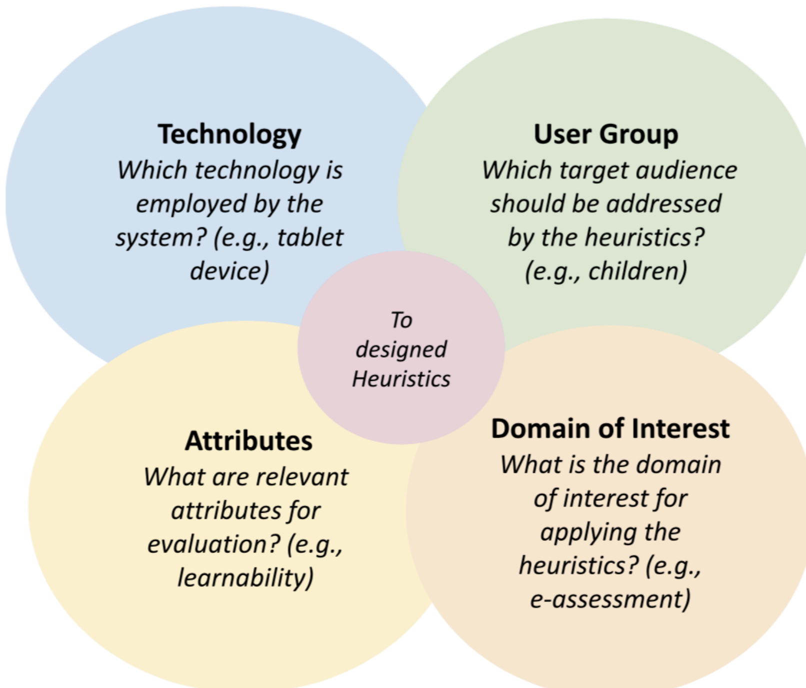


Figure 7. Recommended building blocks for the analysis in the first step of the QRR methodology, the exploratory stage.

Third, in Step 4 of the QRR methodology, we needed help defining concrete e-assessment features. In the examples the authors provided, they did not clarify how they extracted the features (Quiñones et al., 2018). Therefore, we chose to apply an activity diagram to one use case (i.e., an e-assessment application to which we had access). By doing so, we extracted a basic set of e-assessment features. Fourth, Step 5 needed more specifications. In our example, we formulated some exclusion criteria to select and construct a final corpus of guidelines on which to build our heuristics. We decided to synthesize the information by applying thematic analysis, which is different from the method the QRR methodology proposed.

Consequently, there is a need to define this qualitative process in more detail, either by referring to a suitable method or by describing the adjustments. Fifth, we needed help with the process of validating the heuristics. As described in (Quiñones et al., 2018), the expert judgments also contained elements from a heuristic evaluation workshop (i.e., in which the experts were asked to apply the designed heuristics to an application). We constructed our initial survey similarly and ran into the problem that only some participants completed it. We concluded that the survey needed to be shorter (i.e., we asked the participants to relate every problem they found in their e-assessment application to one of our heuristics). In addition to this demanding task, we asked the experts to rate the heuristics on four dimensions (i.e., similarly to what we did on the final survey) and to give advice for improving them. Therefore, we recommend clearly distinguishing between these two methods. An expert judgment survey should collect knowledge about the topic of investigation (i.e., in our case, the triangulation of e-assessments, children, and usability/UX) to validate whether everything from the perspective of multiple experts has been covered. Also, a large pool of experts should check the heuristics for clarity (i.e., to ensure that the heuristics are interpreted correctly).

On the other hand, a heuristic evaluation workshop should ideally focus on the usability aspects of the heuristics (i.e., how does a researcher apply the heuristics, how predictive are the heuristics, and did they detect or cover all problems encountered with a system). The two methods are complementary, and provide the researcher with different data, and could be used at different stages in the development process. Lastly, more explicit guidance is needed to design the surveys and workshops that are being used to validate the heuristics. In this paper, we provided one detailed example of a survey that was used to validate the designed heuristics; however, more work is needed to standardize the tools and measures.

6. Limitations

The current study has several limitations, which we identify in the following. First, the heuristics data belonged to three sources (Gelderblom, 2008; Liu et al., 2022; Sim et al., 2008). To account for the completeness and timeliness of our heuristics, we extracted recent expert feedback with our survey. However, we still might have missed essential evaluation aspects in the current heuristics, even though we had taken numerous experts by the hand (N = 24). Furthermore, given that English is the predominant language in the scientific community, we limited our scope to existing heuristics presented exclusively in English (there was one exception from the German literature, but we did not include it in the final selection). Therefore, we might have missed the community-specific guidelines or heuristics published in languages other than English. However, with the United States and the UK as top countries for publications in the domain of e-assessment (Ortiz-López et al., 2023) and similarly for the other domains (children and usability/UX), we believe that we covered the most important aspects. Second, in the construction of our initial corpus of guidelines from which we extracted the heuristics, we might have introduced some bias, as most of the coding was done by the first author. However, by including two more authors for a cross-check on the exclusions and codes (i.e., thematic analysis) to attain *intersubjective consensus* (Miles et al., 1994), we helped to further reduce bias in the interpretation and coding. Third, we based the features that were extracted for the e-assessment application on one single use case. We could have constructed a similar analysis (e.g., activity diagram) to have a more solid collection of features and to account for a more complete feature set. Finally, we did not use any heuristic to calculate, for example, the effectiveness of our heuristics. It is critical in an early stage to first validate whether the heuristics that were developed were clearly formulated to check for their completeness and features before accounting for their reliability. Additionally, as our heuristics are domain-specific and evidence-based, we do not expect to find fewer usability problems than with a standard set of heuristics (Nielsen, 1992, 1994) (i.e., see also a previous discussion on this topic in Section 5). However, future research should investigate applying the heuristics to different e-assessment systems with different levels of expertise to test and guarantee their transferability. For this purpose, heuristic evaluation workshops would be ideally suited where different systems could be taken by hand to test for the validity and reliability of the heuristics. Additionally, future research could let some experts use the heuristics for extended periods in time (i.e., longitudinal study) to encounter their practical necessity.

7. Conclusion

In this paper, we followed the QRR methodology to design a set of usability/UX heuristics to evaluate children's e-assessments. The proposed heuristics were based on an evidence-based corpus of 506 guidelines, which we further synthesized using the thematic analysis approach. The final set of 10 heuristics was validated through two rounds of expert evaluations: a heuristic evaluation workshop and an online expert survey. The 10 derived heuristics were evaluated as strongly positive in utility, relevance, clarity, and simplicity by relevant domain experts. On the basis of our analysis, we concluded that the existing heuristics provide a foundation for evaluating e-assessment systems for children between the ages of 6 and 12. Future research should invest in validating the heuristics further, for instance by applying them in an heuristic evaluation workshop.

Moreover we put several recommendations in place for the QRR methodology, emphasizing the comprehensiveness of the exploration criteria (i.e., Step 1 of the approach). Hence, we advocate including the user group and technology along with the domain and usability/UX focus. Future research should verify and complete this notion by applying and advancing the current methodology. Considering the points above, we propose a first set of 10 UX/usability heuristics to evaluate the UX and usability of e-assessment platforms designed for children between 6 and 12 years.

8. Acknowledgments

The authors thank the Human-Computer Interaction research group from the University of Luxembourg for providing feedback on the methodology. In particular, we thank Carine Lallemand for helping conceptualize the research, Vincent Koenig for his supervision at the beginning of the study and his feedback on the first draft report, and Jane Zargoski for her professional language and grammar check. We also thank Pedro Cardoso-Leite for his support on the data visualization. Additionally, we would like to thank the “Direction de l'évaluation, de la prospective, et de la performance” for providing us with the necessary research resources (e.g., access to the e-assessment application).

9. References Chapter 4

About PISA. (n.d.). [OECD]. Retrieved December 27, 2023 from https://www.oecd.org/pisa/aboutpisa/

Ager, R., & Kendall, M. (2003). *Getting it right from the start: A case study of the development of a Foundation Stage Learning and ICT Strategy in Northamptonshire, UK*. 3–11.

Al-Azawei, A., Baiee, W. R., & Mohammed, M. A. (2019). Learners' experience towards e-assessment tools: A comparative study on virtual reality and moodle quiz. *International Journal of Emerging Technologies in Learning (Online)*, *14*(5), 34.

Alcaraz Martínez, R., Turró, M. R., & Granollers Saltiveri, T. (2022). Methodology for heuristic evaluation of the accessibility of statistical charts for people with low vision and color vision deficiency. *Universal Access in the Information Society*, *21*(4), 863–894.

Al-Emran, M., Mezhuyev, V., & Kamaludin, A. (2018). Technology Acceptance Model in M-learning context: A systematic review. *Computers & Education*, *125*, 389–412.

Alsumait, A., & Al-Osaimi, A. (2009). *Usability heuristics evaluation for child e-learning applications*. iiWAS '09: Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services, 425–430. https://doi.org/10.1145/1806338.1806417

Arhippainen, L. (2013). *A tutorial of ten user experience heuristics*. AcademicMindTrek '13: Proceedings of International Conference on Making Sense of Converging Media, 336–337. https://doi.org/10.1145/2523429.2523491

Avouris, N. M., Dimitracopoulou, A., Daskalaki, S., & Tselios, N. K. (2001). Evaluation of distance-learning environments: Impact of usability on student performance. *International Journal of Educational Telecommunications*, *7*(4), 355–378.

Bastien, J. C., & Scapin, D. L. (1995). Evaluating a user interface with ergonomic criteria. *International Journal of Human-Computer Interaction*, *7*(2), 105–121.

Bunderson, C. V., Inouye, D. K., & Olsen, J. B. (1988). The four generations of computerized educational measurement. *ETS Research Report Series*, *1988*(1), i–148.

Burmester, M., Hassenzahl, M., & Koller, F. (2002). Usability ist nicht alles–Wege zu attraktiven Produkten (Beyond Usability–Appeal of interactive Products). *I-Com*, *1*(1), 32–40.

Chen, C. J., Toh, S. C., & Ismail, W. M. F. W. (2005). Are learning styles relevant to virtual reality? *Journal of Research on Technology in Education*, *38*(2), 123–141.

Chiasson, S., & Gutwin, C. (2005). Design principles for children's technology. *Interfaces*, *7*(28), 1–9.

Colombo, L., & Pasch, M. (2012). Heuristics for an Optimal User Experience. *Proc. CHI2012 Altchi. ACM Press*.

Diefenbach, S., Kolb, N., & Hassenzahl, M. (2014). *The 'hedonic' in human-computer interaction: History, contributions, and future research directions*. DIS '14: Proceedings of the 2014 conference on Designing interactive systems (pp. 305–314). https://doi.org/10.1145/2598510.2598549

Druin, A. (1999a). Cooperative inquiry: Developing new technologies for children with children. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems the CHI Is the Limit - CHI '99* (pp. 592–599). https://doi.org/10.1145/302979.303166

Druin, A. (Ed.). (1999b). *The design of children's technology*. Morgan Kaufmann Publishers.

Endsley, T. C., Sprehn, K. A., Brill, R. M., Ryan, K. J., Vincent, E. C., & Martin, J. M. (2017). *Augmented reality design heuristics: Designing for dynamic interactions*. *61*(1), 2100–2104.

Franklin, F. F. (2014). *Heurísticas de usabilidade para sistemas colaborativos remotos de realidade aumentada*. In Proceedings of XVI Symposium on Virtual and Augmented Reality, 53–62.

Gabas, C., Cutler, L., & Schachter, R. E. (2023). Making Mistakes: Children's Errors as Opportunities for Emergent Literacy Learning in Early Childhood. *The Reading Teacher*, *76*(6), 664–672.

Gelderblom, J. H. (2008). *Designing Technology for Young Children: Guidelines Grounded in a Literature Investigation on Child Development and Children's Technology.* [Doctoral thesis, Pretoria, South Africa: University of South Africa].

Graham, D. (2004). A survey of assessment methods employed in UK higher education programmes for HCI courses. *Proceedings of the 7th HCI Educators Workshop* (Preston, LTSN), 66–69.

Guest, G., MacQueen, K. M., & Namey, E. E. (2011). *Applied thematic analysis*. Thousand Oaks, CA: SAGE Publications, Inc; 2012:978-971.

Hanna, L., Risden, K., & Alexander, K. (1997). Guidelines for usability testing with children. *Interactions*, *4*(5), 9–14.

Hassenzahl, M. (2001). The effect of perceived hedonic quality on product appealingness. *International Journal of Human-Computer Interaction*, *13*(4), 481–499.

Hassenzahl, M., & Tractinsky, N. (2006). User experience-a research agenda. *Behaviour & Information Technology*, *25*(2), 91–97.

Henderson, S. D., Many, J. E., Wellborn, H., & Ward, J. (2002). How scaffolding nurtures the development of young children's literacy repertoire: Insiders' and outsiders' collaborative understandings. *Literacy Research and Instruction*, *41*(4), 309–330.

Hermawati, S., & Lawson, G. (2015). *A user-centric methodology to establish usability heuristics for* *specific domains*. In Proceedings of the International Conference on Ergonomics & Human Factors, 80-85. https://doi.org/10.1201/b18293

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75–105. JSTOR. https://doi.org/10.2307/25148625

Hornby, W. (2003). Assessing using grade-related criteria: A single currency for universities? *Assessment & Evaluation in Higher Education*, *28*(4), 435–454.

Hourcade, J. P. (2015). *Child-Computer Interaction*. CreateSpace Independent Publishing Platform.

Hub, M., & Čapková, V. (2010). *Heuristic evaluation of usability of public administration portal', DEO, Narsingh, et al*. Applied Computer Science in: Proceedings of International Conference on Applied Computer Science, ACS.

International Organization for Standardization (ISO). (2018). *ISO 9241-11:2018: Ergonomics of human-system interaction—Part 11: Usability: Definitions and concepts.* International Organization for Standardization. Retrieved 8 February, 2024 from https://www.iso.org/standard/63500.html

International Test Commission. (2006). International guidelines on computer-based and internet-delivered testing. *International Journal of Testing*, *6*(2), 143–171.

Jahnke, I., Riedel, N., Singh, K., & Moore, J. (2021). Advancing Sociotechnical-Pedagogical Heuristics for the Usability Evaluation of Online Courses for Adult Learners. *Online Learning*, *25*(4), 416–439.

Jiménez, C., Cid, H. A., & Figueroa, I. (2017). PROMETHEUS: procedural methodology for developing heuristics of usability. *IEEE Latin America Transactions*, *15*(3), 541–549.

Knight, P. (2001). *A briefing on key concepts: Formative and summative, criterion and norm-referenced assessment*. York: LTSN Generic Centre.

Lallemand, C., & Gronier, G. (2018). *Méthodes de design UX. 30 méthodes fondamentales pour concevoir des expériences optimales.* (2nd ed.). Groupe Eyrolles.

Lavery, D., Cockton, G., & Atkinson, M. P. (1997). Comparison of evaluation methods using structured usability problem reports. *Behaviour & Information Technology*, *16*(4–5), 246–266.

Lechner, B., Fruhling, A. L., Petter, S., & Siy, H. P. (2013). *The Chicken and the Pig: User Involvement in Developing Usability Heuristics.* AMCIS.

Lehnert, F. K. (2019). *User Experience challenges for designing and evaluating Computer-Based Assessments for children*. IDC '19: Proceedings of the 18th ACM International Conference on Interaction Design and Children, 736–741. https://doi.org/10.1145/3311927.3328724

Lehnert, F. K., Niess, J., Lallemand, C., Markopoulos, P., Fischbach, A., & Koenig, V. (2022).

Child–Computer Interaction: From a systematic review towards an integrated understanding of interaction design methods for children. *International Journal of Child-Computer Interaction*, *32*, 100398. https://doi.org/10.1016/j.ijcci.2021.100398

Lewis, C., & Wharton, C. (1997). Cognitive walkthroughs. In *Handbook of human-computer interaction*, 717–732. Elsevier.

Liebal, J., & Exner, M. (2011). *Usability für Kids*. Springer.

Liu, F., Sherwin, K., & Budiu, R. (2022). *UX Design for Children (Ages 3-12)* (4th ed.). CA: Nielsen and Norman Group.

Loewenberger, P., & Bull, J. (2003). Cost-effectiveness analysis of computer-based assessment. *ALT-J*, *11*(2), 23–45.

Markopoulos, P., & Bekker, M. (2003). On the assessment of usability testing methods for children. *Interacting with Computers*, *15*(2), 227–243. https://doi.org/10.1016/S0953-5438(03)00009-2

Markopoulos, P., Read, J. C., MacFarlane, S., & Hoysniemi, J. (2008). *Evaluating Children's Interactive Products*. Morgan Kaufmann.

Miles, M. B., Hubermann, A. M., & Saldana, J. (1994). *Qualitative Data Analysis: A Methods Sourcebook* (3rd ed.). Sage.

Morville, P. (2004). *User Experience Design*. Retrieved December 27, 2023 from http://semanticstudios.com/user_experience_design/

Nielsen, J. (1992). *Finding usability problems through heuristic evaluation*. In Proc. CHI '92: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 373–380. https://doi.org/10.1145/142750.142834

Nielsen, J. (1994). *Enhancing the explanatory power of usability heuristics*. In Proc. CHI '94: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 152–158. https://doi.org/10.1145/191666.191729

Nielsen, J. (2012). Usability 101: Introduction to Usability. *Usability 101: Introduction to Usability*. Retrieved December 27, 2023 from https://www.nngroup.com/articles/usability-101-introduction-to-usability/

Nielsen, J., & Mack, R. (2014). *Usability Inspection Methods*. John Wiley & Sons.

Nielsen, J., & Molich, R. (1990). *Heuristic evaluation of user interfaces*. In Proc. CHI '90: Proceedings of the SIGCHI Conference on Human Factors in Computing System, 249–256. https://doi.org/10.1145/97243.97281

Ortiz-López, A., Olmos-Migueláñez, S., & Sánchez-Prieto, J. C. (2023). Toward a new educational reality: A mapping review of the role of e-assessment in the new digital context. *Education and Information Technologies*, 1–28.

Tsai, P. & Oppenheimer, D. (2021). *Making grading in university courses more reliable*. Retrieved December 27, 2023 from https://www.timeshighereducation.com/campus/making-grading-university-courses-more-reliable

Parshall, C. G., Davey, T., & Pashley, P. J. (2000). Innovative item types for computerized testing. *Computerized Adaptive Testing: Theory and Practice*, 129–148.

Piaget, J. (1955). *The Construction of Reality in the Child*. NYBasic Books.

Quiñones, D., & Rusu, C. (2017). How to develop usability heuristics: A systematic literature review. *Computer Standards & Interfaces*, *53*, 89–122.

Quiñones, D., Rusu, C., Arancibia, D., González, S., & Saavedra, M. J. (2020). SNUXH: A set of social network user experience heuristics. *Applied Sciences*, *10*(18), 6547.

Quiñones, D., Rusu, C., & Rusu, V. (2018). A methodology to develop usability/user experience heuristics. *Computer Standards & Interfaces*, *59*, 109–129.

Redecker, C., & Johannessen, Ø. (2013). Changing assessment—Towards a new assessment paradigm using ICT. *European Journal of Education*, *48*(1), 79–96.

Riley, S. K. L., & Stacy, K. (2008). Teaching in virtual worlds: Opportunities and challenges. *Setting Knowledge Free: The Journal of Issues in Informing Science and Information Technology*, *5*(5), 127–135.

Romeo, G., Edwards, S., McNamara, S., Walker, I., & Ziguras, C. (2003). Touching the screen: Issues related to the use of touchscreen technology in early childhood education. *British Journal of Educational Technology*, *34*(3), 329–339.

Rowntree, D. (2015). *Assessing students: How shall we know them?* Routledge.

Rusu, C., Roncagliolo, S., Rusu, V., & Collazos, C. (2011). *A Methodology to establish usability heuristics*. In Proc. of ACHI 2011 : The Fourth International Conference on Advances in Computer-Human Interactions, 59–62.

Scaife, M., & Bond, R. (1991). Developmental changes in childrens' use of computer input devices. *Early Child Development and Care*, *69*(1), 19–38.

Shneiderman. (1998). *Designing the User Interface* (3rd edited.). Addison Wesley Longman.

Siddiq, F., Gochyyev, P., & Wilson, M. (2017). Learning in Digital Networks–ICT literacy: A novel assessment of students' 21st century skills. *Computers & Education*, *109*, 11–37.

Sim, G. (2009). *Evidence based design of heuristics: Usability and computer assisted assessment*. [Doctoral thesis: University of Central Lancashire].

Sim, G., & Read, J. C. (2016). Using computer-assisted assessment heuristics for usability evaluations. *British Journal of Educational Technology*, *47*(4), 694–709.

Sim, G., Read, J. C., & Holifield, P. (2008). *Heuristics for evaluating the usability of CAA applications*. IN: Khandia, F. (ed.). 12th CAA International Computer Assisted Assessment Conference: Proceedings of the Conference on 8th and 9th July 2008 at Loughborough University. Loughborough : Loughborough University, 283-294. https://hdl.handle.net/2134/5545

Sim, G., Read, J. C., Holifield, P., & Brown, M. (2007). *Heuristic Evaluations of Computer Assisted Assessment Environments*. In Proc. EdMedia: World Conference on Educational Media and Technology. Association for the Advancement of Computing in Education (AACE), 896-905.

Squires, D., & Preece, J. (1999). Predicting quality in educational software. *Interacting with Computers*, *11*(5), 467–483.

Taras, M. (2002). Using assessment for learning and learning from assessment. *Assessment & Evaluation in Higher Education*, *27*(6), 501–510.

Tselios, N., Avouris, N., & Komis, V. (2008). The effective combination of hybrid usability methods in evaluating educational applications of ICT: Issues and challenges. *Education and Information Technologies*, *13*, 55–76.

Van Greunen, D., Yeratziotis, A., & Pottas, D. (2011). *A three-phase process to develop heuristics*. In Proc. of the 13th Annual Conference on WWW Applications, Johannesburg (2011)(pp. 5–23).

Weinerth, K. (2015). *How does usability improve computer-based knowledge assessment?* [Doctoral thesis: University of Luxembourg].

Weinschenk, S., & T. Barker, D. (2000). *Designing Effective Speech Interfaces*. Wiley.

Wongvorachan, T., Bulut, O., Tsai, Y.-S., & Lindner, M. A. (2022). Improving Student Feedback Literacy in e-Assessments: A Framework for the Higher Education Context. *Trends in Higher Education*, *1*(1), 16–29.

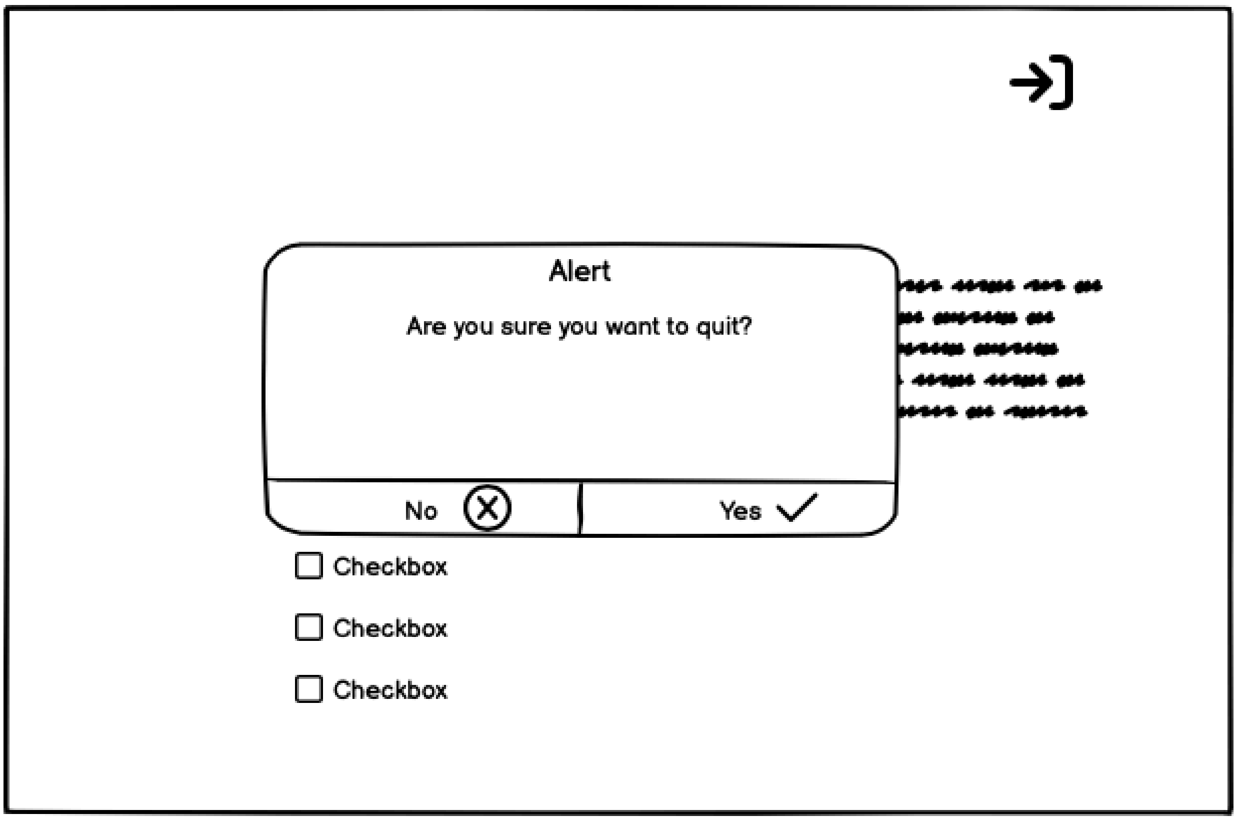
10. Appendices

A. Overview of heuristics

#1 Prevent errors and help to recover from them
#2 Ensure timely and visual feedback on children's actions
#3 Provide support and guidance throughout the interface
#4 Make children's actions and system status transparent
#5 Give children freedom to control the interface
#6 Anticipate consistency across the interface
#7 Design age appropriate interface elements
#8 Make the user interface intuitive and accessible for all children
#9 Account for diverse technology literacy
#10 Eliminate disruptive and external factors of the delivery platform

#1 Prevent errors and help to recover from them

Explanation: Prevent errors with an interface from affecting test performance. If errors occur, help children to recover from their mistakes.
Example: Avoid children from quitting the application by mistake by, e.g., providing a pop-up window in which they need to confirm their action before leaving the application..



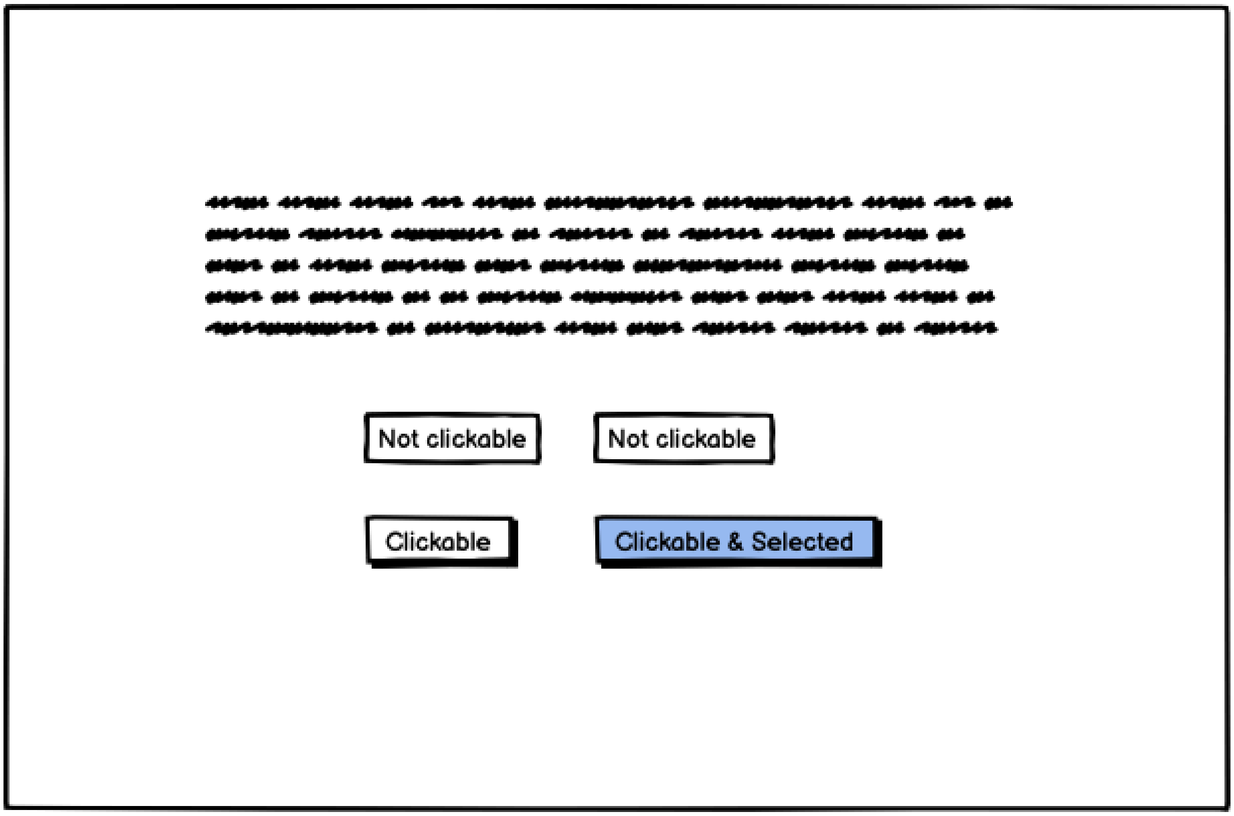
Checklist:

- To prevent children from producing inaccurate outcomes, include restrictions limiting activities they may perform at certain stages throughout the interaction as error prevention measures.
- Assist children in identifying mistakes by providing error messages in languages that children can comprehend. Clearly explain the problem and propose a solution. Make it necessary for children to confirm actions that may be incorrect before executing them, such as marking exits and warning children when they are about to leave the platform.
- Facilitate the process of undoing an activity. Assist in error recovery by providing appropriate guidance based on the individual's comprehension level (e.g., young children may not possess the knowledge to utilize Undo/Redo commands).
- Ensure constant accessibility of a keyboard, whether physical or digital, to correct errors that arise from handwriting interfaces, including spelling problems, grammatical faults (such as improperly formed letters), or execution errors (such as failure to make contact with the tablet using the pen or unintentional addition of characters).

Benefit: Children can concentrate on the assessment and don't use their cognitive capacity for error handling.
E-assessment feature: *Immediate Feedback*
Usability attribute: *Effectiveness, Errors, Learnability*
UX attribute: *Accessibility*
Related heuristics: Gelderblom: F.1.29, F.7.28, F.7.29, F.7.30, F.7.31, Liu et al.: 13, 28, 97, 141, 146, 147, Sim & Read: 1, 7

#2 Ensure timely and visual feedback on children's actions

Explanation: Providing adequate and immediate visual feedback on children's actions and showing them the effect it had on the screen.
Example: Highlight or change the color of a button immediately after it has been pressed.



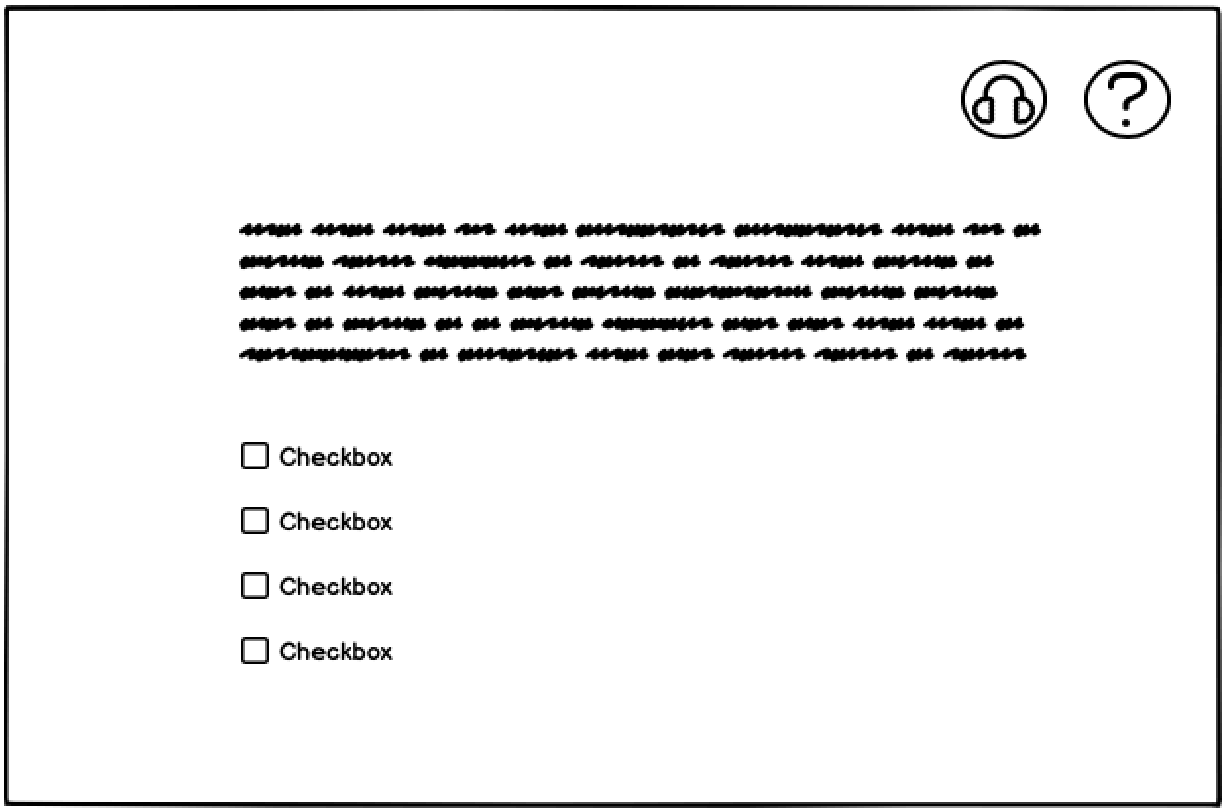
Checklist:

- Offer comprehensive feedback by providing information through several modalities (i.e., acoustic, tactile, verbal, or visual) on the user's actions and corresponding outcomes.
- Present the system's features using visually oriented interface elements instead of relying on text, allowing young children who cannot yet read to utilize the system. Ensure that clickable elements are visually distinguishable as interactive; create buttons, links, and symbols that provide clear cues about their functionality;
- Modify the visual appearance of clickable items.
- Differentiate between navigation and content. Differentiate between objects that youngsters can engage with and those that they cannot;
- Ensure that the response time is prompt. When the job is not completed immediately, the system should signal that it is still being processed. The absence of immediate feedback results in repetitive clicking or keystrokes, potentially affecting the program's operation. Response times for similar jobs should be equivalent;
- Create multimedia content tailored to your target audience's computer capabilities. Provide prompt feedback and refrain from utilizing automated responses for video and animation.

Benefit: Clear navigation structure and faster reaction times because children understand intuitively where they are supposed to click. Also fewer user errors and frustrations.
E-assessment feature: *Immediate feedback, interactive elements*
Usability attribute: *Efficiency*
UX attribute: -
Related heuristics: Gelderblom: F.7.20., F.7.23., C.2.2, C.2.3, C.2.4, C.2.5, C.2.9, C.2.14. Liu et al.: 16, 17, 20, 21, 22, 23, 27, 39, 40, 43, 102, 140, 150, Sim & Read: 3

#3 Provide support and guidance throughout the interface

Explanation: Especially children profit from voluntary support. Scaffolding options should help the child navigating throughout the interface;
Example: Place easy-to find links to brief text/audio explanations wherever the child might or want instructions.



Checklist:

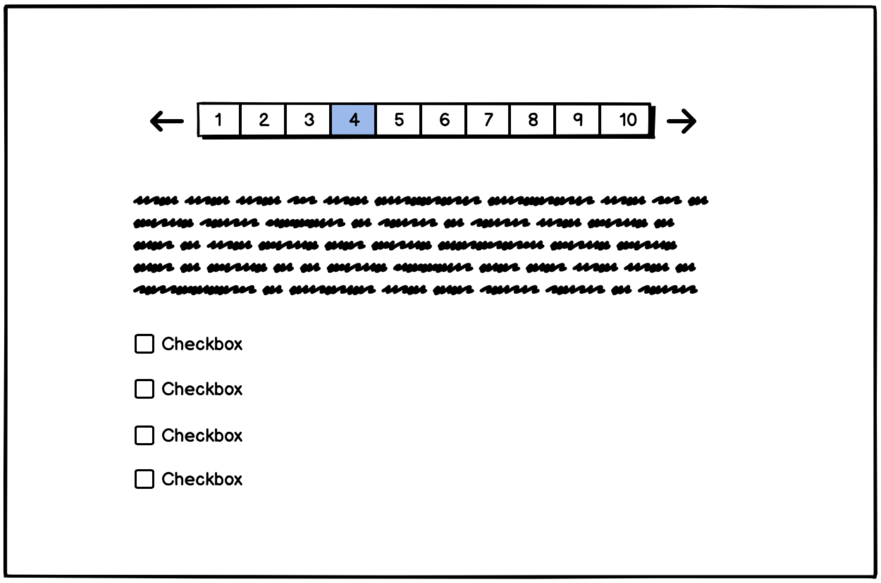
- It is not advisable to rely on trial and error for children to solve all difficulties. Incorporate techniques to direct their attention and prompt them to strategize their answers;
- Ensure that instructions are consistently available. Inform children about the many available scaffolding alternatives and provide instructions on how to access them.
- Include augmentative features such as text-to-speech technology, particularly with a headset. Do not mandate that minors obtain assistance through documents.
- Facilitating the child's attention towards the appropriate interface element with techniques such as highlighting, emphasizing, organizing, and arranging;

Benefit: Teachers and/or Test administrators are not overwhelmed by providing support and children can solve their problems independently. This can have a positive effect on the user experienceUX with the e-assessment platform.
E-assessment feature: *Immediate feedback, independence, interactive elements, monitoring*
Usability attribute: *Learnability, effectiveness*
UX attribute: *Accessibility*

Related heuristics: Gelderblom: B.2.15, B.5.5, B.5.12, C.1.2, C.1.3, C.1.8, C.1.11; Liu et al.: C.1.11; Sim & Read: 3

#4 Make children's actions and system status transparent

Explanation: The children should get insights into the navigation process and system status.
Example: Progress indicators enable children to locate where they have been, where they are and where they can go.



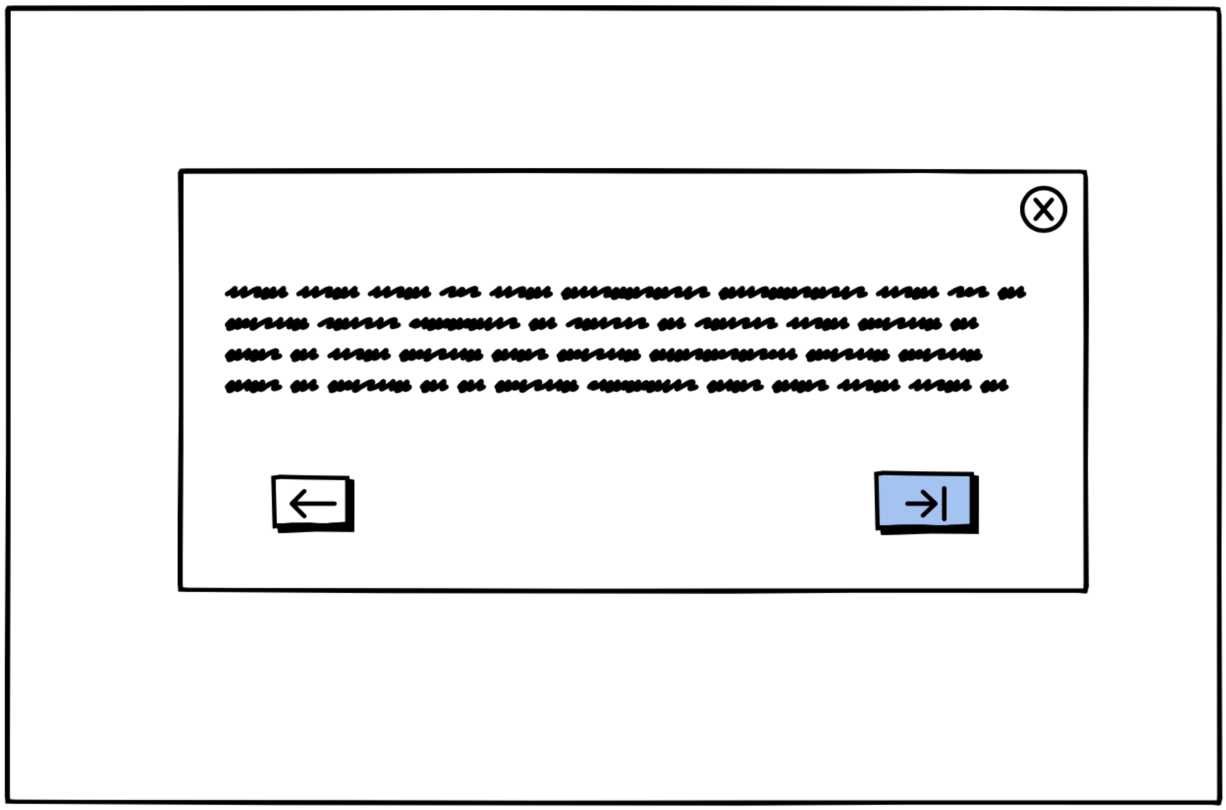
Checklist:

- Render things and activities perceptible to obviate the necessity for children to recollect instructions or prior selections; Ensure that children have a precise understanding of the subsequent necessary course of action. Develop a translucent interface that allows children to concentrate on the essential tasks rather than the mechanics of operating the interface. There must be clear connections between user actions and the resulting impact of these activities (see heuristic #2, #7).
- Refrain from displaying limited navigation cues and prompting visitors to click for further options. To clarify, include all assessment questions in the overview, preferably with corresponding numbers.
- Provide clear and conspicuous "you are here" responses to the children. Present them with their current location and available destinations. Ensure clarity in navigating the program using features such as progress indicators or system status highlighting (see heuristic #2).

Benefit: Transparent user actions help the children to understand the progress and underlying status of their actions so that errors can be prevented such as repeated clicking because the child thinks that the input was not received.
Related heuristics: Gelderblom: F.4.3, F.7.18, F.7.21, F.7.22 Liu et al.: 45, 55, 76, 80, 117, 144; Sim & Read: 4, 9
E-assessment feature: *Independence, monitoring*
Usability attribute: *Memorability*
UX attribute: *Accessibility*

#5 Give children freedom to control the interface

Explanation: Give children the most flexibility for controlling the interface.
Example: Let them freely move within the interface e.g., by reviewing and editing content or skipping instructions.



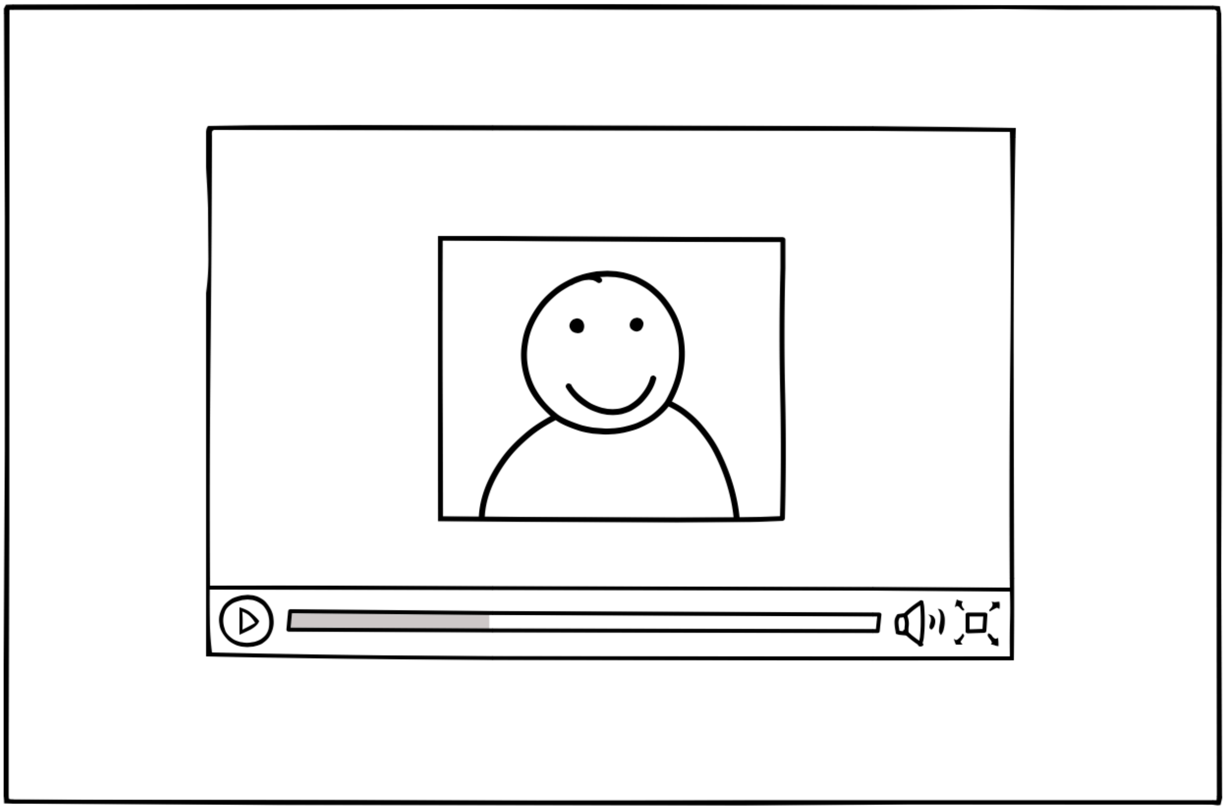
Checklist:

- Ensure that children can modify their activities to observe the resulting impact. To facilitate children's understanding of cause and effect relationships, it is beneficial to allow younger children to manipulate items on the screen by dragging them.
- Enable the user to manipulate multimedia clips as long as it does not limit the assessment objective. Implement a commencement, termination, and repetition feature for all videos.
- Facilitate the utilization of shortcuts for regular users, allowing them to bypass introductions and instructions that they are already familiar with. Ensure that children are knowledgeable of these alternatives;

Benefit: More user control can prevent frustrations (e.g. skipping an introduction) and promote engagement and self-directed exploration of the interface.
E-assessment feature: *independence*
Usability attribute: -
UX attribute: *Accessibility*
Related heuristics: Gelderblom: B.2.3, Liu et al.: 124, E.2.3, F.2.2, F.2.3, F.3.7; Sim & Read: 2

#6 Anticipate consistency across the interface

Explanation: We should strive towards consistent interface design within and across the platform.
Example: An icon (e.g. play button) should represent the same meaning across other platforms used by children.



Checklist:

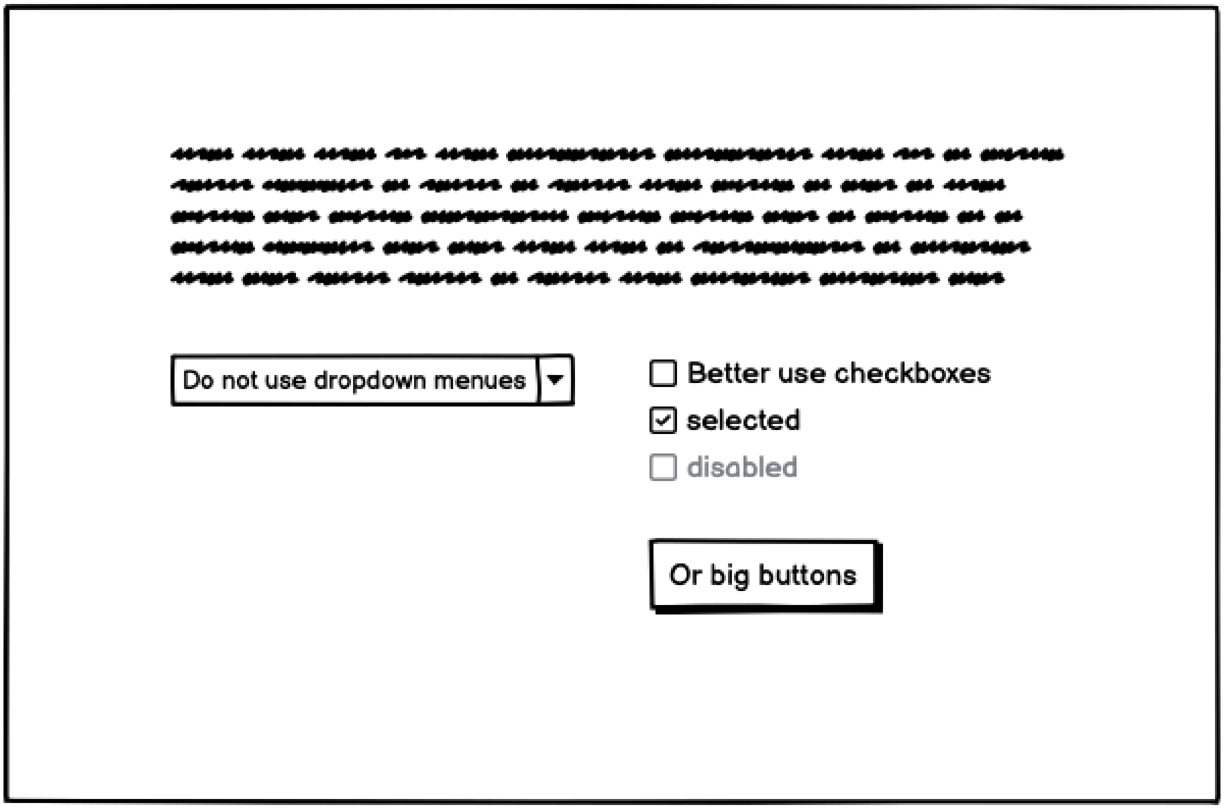
- Use consistent buttons or actions throughout the interface.
- Ensure the coherence and uniformity of the interface and the way information is processed both inside a single system and across several systems. Children anticipate consistent behavior from the system upon repeating an activity. Hence, it is imperative to ensure the activities are analogous.
- Children should be able to match real-world experience with the interface elements, e.g., instead of the Undo/Redo option, you could provide them with an Eraser tool.

Benefit: Age appropriate interface design prevents users from making errors and elicit frustrations. Good interface design can lead towards better UX and fair assessment chances.
E-assessment feature: *maintaining standards*
Usability attribute: *Learnability, effectiveness*
UX attribute: *Accessibility*

Related heuristics: Gelderblom: F.1.11, F.1.27, F.1.28., F.6.8., F.7.1., F.7.4., F.7.5., F.7.20., F.7.23, Liu et al.: 9, 20, 21, 22, 23, 27, 32, 39, 43, 49, 52, 70, 71, 72, 73, 74, 75, 77, 79, 91, 93, 102, 113, 122, 123, 125, 127, 140, 150, 156, Sim & Read: 5

#7 Design age appropriate interface elements

Explanation: Make sure that the interface matches standards in designing for children and addresses age appropriate motor coordinations for that age group.
Example: Make a button big enough so that the child can select it.



Checklist:

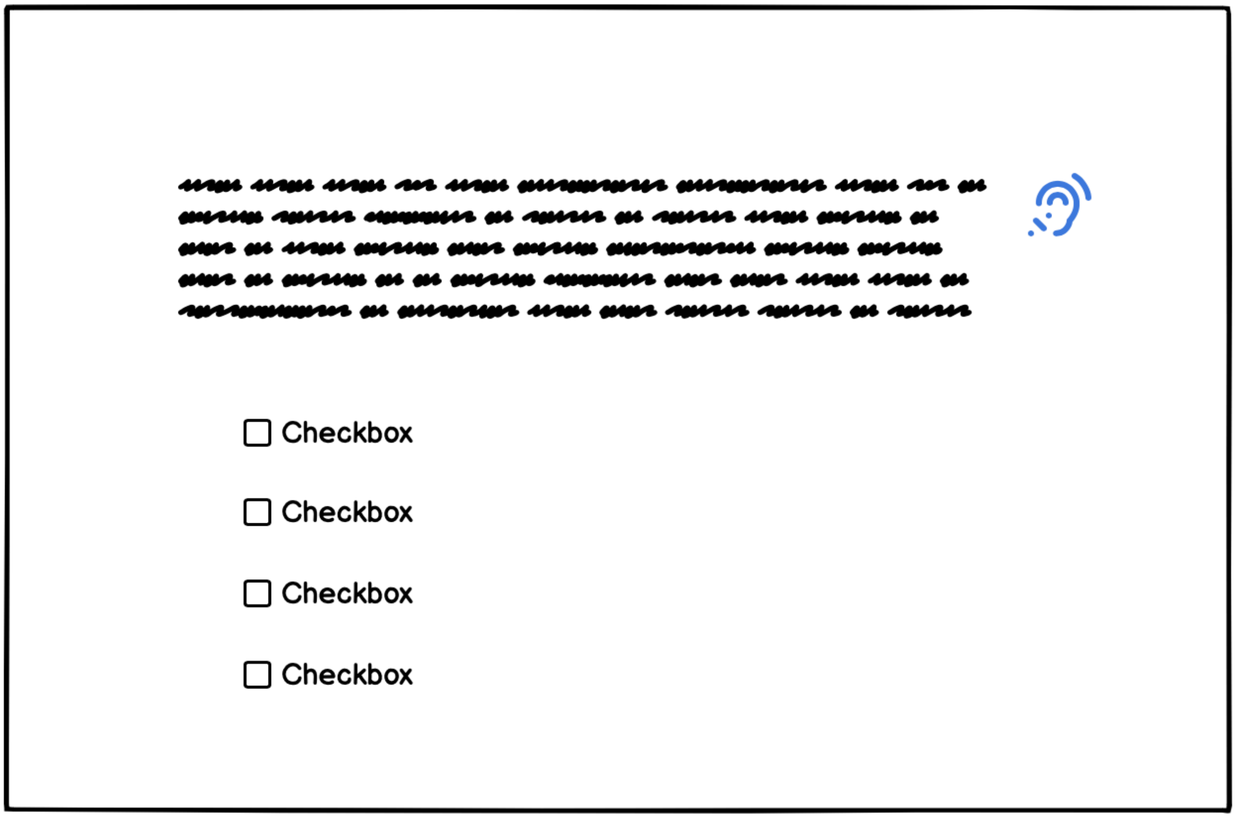
- Ensure that interactions align with children's precise motor skills and developmental levels. Children aged six and older can use input devices effortlessly and utilize many input devices.
- Equip novice writers with a digital pen since it enhances the ability to write smoothly and more effectively than a keyboard.
- **Dimensions of interface components:** Enlarge screen objects when a touchscreen interface is employed, as it is more challenging to select items with a finger accurately. For instance, ensure that the close button is sufficiently large to be easily seen by children.
- **Designing icons:** The accurate interpretation of an icon relies on its caption (the intended message it conveys), the context in which it will be displayed, and the picture itself. When creating icons for children's interfaces, it is essential to consider that the design should utilize children's preexisting knowledge, which may include information acquired from watching cartoons (see Heuristic #6).
    - Children prefer colored icons and perceive black-and-white symbols as more challenging to identify.
    - They have a preference for icons that are presented in a boxed format;

    - Children have greater ease in identifying animated symbols than static ones, and they derive pleasure from the icons becoming animated upon the movement of the mouse over them.
    - Icons that include verbal clues are inappropriate for children between five and eight.
    - Cultural specificity should be avoided while designing icons.
- **Text design:**
    - Use simple, relatively large fonts, comparable in size to at least 12-point print type;
    - If you have to use text, use easily understandable and succinct text only;
    - Minimize the amount of text on screens. Chunk text into short paragraphs for easier reading and scanning;
    - Combine text with images to encourage reading; use audio, pictures, and animations to supplement instructional text;
    - Do not write text in upper case only;
    - Provide good contrast between the text and the background, place text on solid backgrounds;
    - Avoid presenting documents as PDFs. Have HTML versions available;

Benefit: Prevent children from making errors or not being able to navigate throughout the interface.
E-assessment feature: *maintaining standards*
Usability attribute: *learnability, effectiveness*
UX attribute: *Accessibility*
Related heuristics: Gelderblom: A.2.2., F.6.9., F.1.13., F.1.25., F.7.3., F.7.7., F.7.23. Liu et al.: 8, 24, 25, 48, 56, 102, 116, 126, Sim & Read: 4

#8 Make the user interface intuitive and accessible for all children

Explanation: Make sure the interface is accessible to all children and they can adjust it to their needs.
Example: The interface should be supported by auditive input.



Checklist:

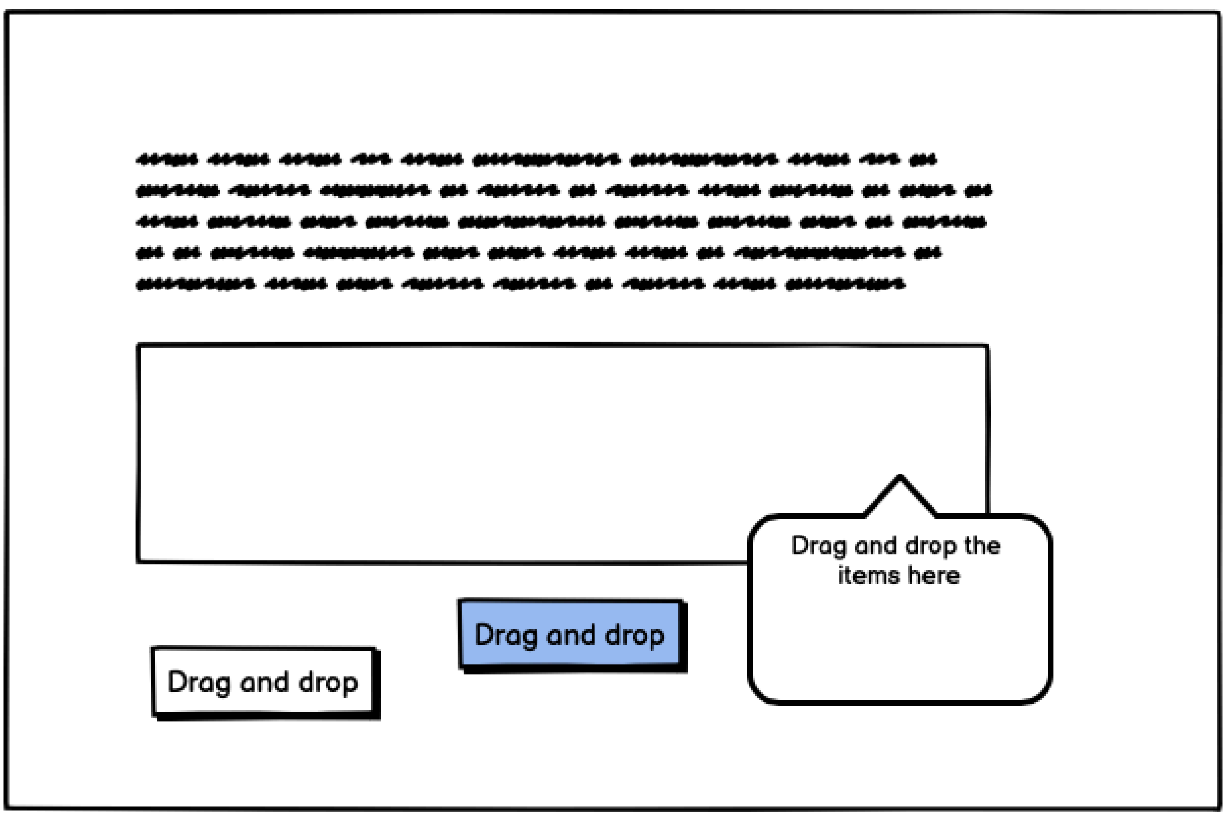
- Utilize multi-modal feedback to enhance the clarity and usability of technology designed for children. Offer audio instructions, ensuring negligible disruption to other children by requiring headphones. For example, audio and text cues can enable children who can read and those who are learning to access a system. Additionally, it may aid in ensuring that a plan is accessible to children with impairments;
- The user interface should possess adaptability, enabling users to modify the input and output format as desired. For instance, the interface should allow users to turn noises or music on or off.
- Adaptive systems provide automated user interface customization, such as allowing children in multilingual nations to select their favorite language.
- Ensure that children possess a precise comprehension of numbers and ideas that bear phonetic resemblance to other words (e.g., two, table, odd). Always take into account how the context could impact their comprehension;

Benefit: Equal chances for all users of the the same interface due to improved accessibility of the interface;

E-assessment feature: *Interactive elements, Independence*
Usability attribute: *learnability, effectiveness*
UX attribute: *Accessibility*
Related heuristics: Gelderblom: F.7.33; F.7.34; Sim: 8

#9 Account for diverse technology literacy

Explanation: Children may not have the same technology literacy so we have to prevent them from making errors based on fewer experiences with the assessment interface or a particular device such as the tablet.
Example: Including a tutorial page where children can practice certain functionality (e.g. drag and drop) of the interface.



Checklist:

- Do not underestimate the importance of the following computer literacy skills:
    i. Understanding the computer's interface and being knowledgeable about the norms associated with it;
    ii. Interacting with digital text;

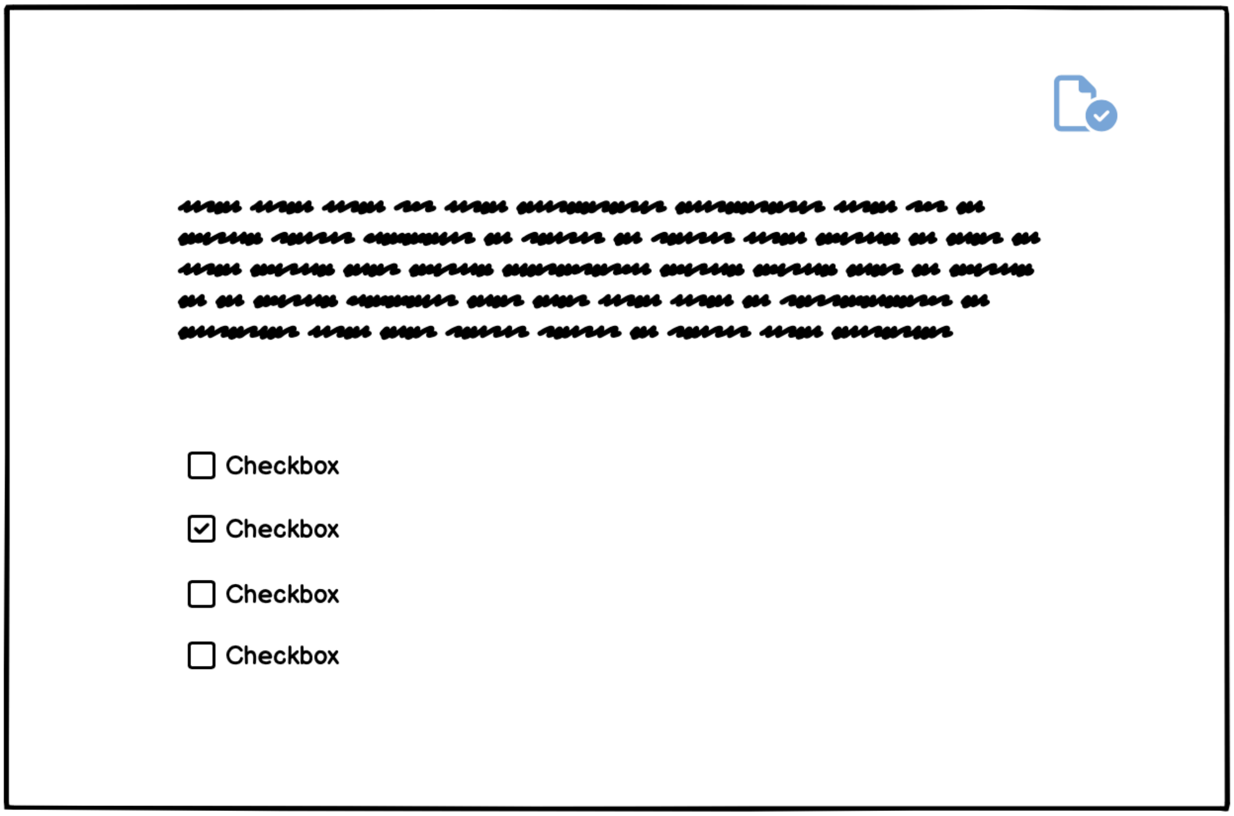
iii. Searching for and obtaining relevant information;
iv. Possessing the requisite sensorimotor abilities for engaging with computers;

- Initiate children with activities that allow them to develop proficiency in utilizing the touchscreen interface. This may be accomplished by incorporating a training module that assists youngsters in acquiring a sufficient cognitive framework to navigate the system effectively. Examples of practice include:
  i. Manipulating screen-based items by dragging and dropping them;
  ii. Identifying the precise positions of each key on the keyboard;

Benefit: Exercising the different interaction functions in the platform before the real assessment starts can free up working memory. So children are not limited to previous knowledge and can concentrate on the task at hand.
E-assessment feature: *Independence*
Usability attribute: *errors, learnability, memorability*
UX attribute: *accessibility*
Related heuristics: Gelderblom: B.6.1, B.6.2, F.1.10, F.7.26, Sim & Read: 10

#10 Eliminate disruptive and external factors of the delivery platform

Explanation: The delivery platform should be responsive, secure and robust.
Example: The answers to questions should be automatically saved.



Checklist:

- Design multimedia for your audience's computing power. Ensure that there is minimal latency when moving between questions or saving answers;
- Ensure the delivery platform is secure and robust, e.g., children are already aware about privacy issue;
- When answers are input, the data should not be lost or overridden by automatic input;

Benefit: Children stay concentrated and do not lose their focus on the e-assessment.
E-assessment feature: *Maintaining standards*
Usability attribute: *efficiency, errors*
UX attribute: -
Related heuristics: Gelderblom: B.6.1, B.6.2, F.1.10, F.7.26, Sim & Read: 10

B. Individual form Evaluators' review

Severity Rating:

1. **Cosmetic Problem:** The issue identified is mainly *aesthetic*. Correcting it will improve the interface's layout and make the tool more pleasant to use; It is not expected to affect the performance with thee-assessment, it would be great to fix if extra time is available.
2. **Minor problem:** The severity of the issue is *medium,* and it can be overcome by the user with some extra effort. Correcting it will improve the usability of the tool and the overall experience; it might impact the test performance of some students (e.g. students with less computer/tablet literacy skills).
3. **Major Problem:** The severity of the issue is *important*, and it is expected to create difficulties for all users regardless of their literacy. Correcting it is expected to have a direct impact on the usability of the tool and the validity of the test; it could affect the results of the test, and fixing it should be given high priority.

Problem found (write a single problem in each space)	Heuristic(s) violated (Insert #n of heuristic)	Task (where it was violated)	How it was found	Severity Rating (1 - 4)	Comments about Heuristic
		Starting the test Navigating within the test Answering a question Finishing the test	Scanning for problems Applying the heuristic to search for problems Trying to force errors		
		Starting the test Navigating within the test Answering a question Finishing the test	Scanning for problems Applying the heuristic to search for problems Trying to force errors		

C. Merged Problems Evaluators

Problem (state the problem here)	Where did it happen?	Frequency	Heuristic violated	Severity score
Entry button is missing an icon; there is no visual feedback; it does not look like a button, no hover state, and no color change.	After the tutorial (entry page)	1	7, 2	1
No icon is presented and there are no audio instructions; thus, students are forced to read. The verification process has to be done by the students.	ID and Password	1	11	1
The arrow to go back to the menu does not look clickable; also there is no button to start the assessment.	Entry page (after selecting the module)	1	7	1
There is no visual feedback, and the sound is turned off (need to change the settings to turn it on); to listen to the text, a button has to be pushed. On another page, the sound started automatically.	Tutorial page	1	3, 2	2
The tutorial is moving automatically to the next page/questions, might be too fast for some kids; Lack of control during onboarding, e.g., students cannot skip or move back to the tutorial.	Tutorial page	1	6, 10	2
There is a lack of explanation, the voice does not mention the verb clicking or press, you do not know how to interact.	Tutorial page	1	3	2

The tutorial is passive, and they do not know where to click, nor can they practice with the interaction elements (e.g., drag and drop)	Tutorial page	1	11	2
The contrast between the background and text is not good, also red and green are used together, which is problematic for children with color-blindness.	Tutorial page	1	7	1
Text is centered and not left-aligned; it should be all aligned similar to the other content.	Tutorial page		7	1
Problems with accessibility, (i.e., it is hard to read), it is not clear how to organize the info, small sentences (white on blue); there is no defined area for the questions, overall not well-structured, and a lack of consistency.	Start test	all time , every page	9, 6	3 (accessibility), 2 (for the structure)
The icon at the top left is for back navigation, it is not intuitive, no other navigation option.	Test page	every page	3, 5	1
Overall lack of consistency, if you don't understand the icon, you can't go back. No clear differentiation between images vs. icons.	Test page	every page	7, 6	2

Lack of transparency regarding the navigation. Students should be warned about the consequences of clicking the “next” or “back” buttons	Test page	every page	4	3
If you use the logo button, for some questions, it deletes your entry data.	Test page	every page	4	2
If there is no data input, there should be a warning that no data have been typed before proceeding.	Test page	every page	4	3
Lack of consistency with well-known icons and their meaning (i.e., arrow means play for a specific part of the text but the arrow is not play for the kids); headphone icon for voice input on the tutorial page (not as expected for listening).	Test page	every page	6	1
The arrow button looks like a dropdown, not like an audio explanation, the call to action is missing.	Test page	every page	7	1
There is no visual feedback after clicking the audio button.	Test page	every page	2	1
Lack of consistency, different color background, same type of question but different way of presenting them.	Test page		6	1
Half of the sentence moved off the screen and is therefore difficult to read, problem with the responsiveness of the page? Also some elements are hidden.	Test page		6 (on technical level), 12, 9	3
No visual feedback and not sure where you can drag and drop.	Test page		3, 5	2

Drag and drop lacks consistency (i.e., different colors).	Test page		6	1
There is a mismatch within one question between what is asked and what can be done, you should select one item, and you have to select two answers.	Test page	1	4, 6	3

Chapter 5: General Discussion

1. Synthesis of results

The primary objective of this doctoral dissertation was to investigate and report on methods, including their challenges, for measuring children's user experience (UX) in an e-assessment context. My coauthors and I first aimed to provide an overview of recent developments and empirical methodologies in child-computer interaction (CCI). By conducting this overview, we identified suitable methods for investigating children's UX with e-assessments. To paint a holistic picture, we examined UX from two angles: the children's self-reported UX and the experts' perspective. We also used two distinct methods: a survey focusing on scale evaluation (children's self-reports) and a heuristic evaluation (experts' perspective).

1.1. Recent developments in measuring children's UX

The first objective of this dissertation was to establish an overview of existing interaction design methodologies employed in the context of e-assessments administered to children. The ultimate corpus included in our systematic literature review in Chapter 2 (Lehnert et al., 2022) comprised a total of 272 empirical publications. Through our investigation, we improved our understanding of the current empirical approaches employed in CCI. The findings suggested a tendency toward implementing field studies, with a particular emphasis on specific product categories, such as interactive games and professional software. In addition, we documented an increase in research efforts related to new technologies, such as virtual reality (VR) and augmented reality (AR), in recent years.

Moreover, a triangulation of methods is commonly applied, with more than half of the researchers employing a mix of two or more approaches in the design and evaluation of their products. However, our findings indicate that the most prevalent approach in research is the combination of various research-led methods (i.e., user observation and semi-structured interviews). Therefore, in future research, we encourage researchers to explore novel combinations of techniques from the broad spectrum of research- and design-led approaches with integrations of different levels of stakeholders during the research process. The problems we identified also highlight the necessity for additional strategies and interventions that are appropriate for younger children and those with various abilities (Spiel et al., 2019). There is a notable endeavor to enhance the range of approaches in the area of participatory design, including the utilization of creative sessions among youngsters. This endeavor is particularly evident in the advancement of novel tools and procedures, as such an expansion is expected to persist into the future and encompass all growth segments. Furthermore, it is essential to highlight the necessity of reporting on research practices (Spiel & Gerling, 2021) by, among others, documenting the use of ethical protocols when engaging children in research and creative endeavors. Taking into consideration the challenges involved in such thorough documentation, we anticipate forthcoming patterns and obstacles in the utilization of CCI techniques to achieve a child-centered approach.

1.2. Challenges involved in measuring children's self-reported UX

The second objective of this dissertation was to evaluate the suitability of survey and scale designs for measuring children's self-reported UX. We selected a survey as the possible methodology for evaluating children's UX because it is a discrete method that does not require researchers to intervene. This technique can be implemented both before and after an evaluation. However, it is not advisable to administer a survey during the assessment, as it could distract the children from the primary activity and could impact their overall performance on the exam. Additionally, surveys offer the benefit that they can be administered to a large sample which was an advantage and privilege we had due to the collaboration with our partners (i.e., DEPP).

Nonetheless, more methodological knowledge is needed on how best to administer well-designed response scales that are appropriate for children's cognitive capacities, levels of motivation, and environmental circumstances. Extreme replies (Read et al., 2002.; Zaman et al., 2012), nonresponse, random answers, low repeatability, and inconsistency are frequently recorded (Borgers, 2003; Borgers et al., 2004; Borgers & Hox, 2001), particularly with young children (J. C. Read & MacFarlane, 2006; Zaman et al., 2012) who are still in elementary school. Thus, Chapter 3 presented an empirical assessment of three often utilized visual scale designs, namely, the Nodding Heads Scale (NHS), the Smileyometer Scale (SOS), and the Thumbs-Up Scale (TUS). The findings shed light on various diverse factors that tend to influence children's answers and potentially bias the survey questions. As previously reported (Read et al., 2002; Zaman et al., 2012), these biases include nonresponse and overly positive responses. In a separate academic paper that is related to but not included in the present dissertation, we examined the factors that have the potential to contribute to bias by explicitly focusing on response and anchor biases observed in survey research.

In Chapter 2 and the additional research results, we demonstrated how important it is to develop scales that are pleasant but simultaneously less complicated (i.e., with fewer answer alternatives and no ambiguity), as these characteristics may affect children's responses to survey questions. To illustrate, in our cognitive interview study, the NHS was rated as lowest on enjoyment and the highest in terms of complexity while at the same time showing the highest level of nonresponse in the field research (i.e., digital questionnaire). On the other hand, the SOS contained a higher completion rate (i.e., digital questionnaire) and was evaluated to have a lower level of difficulty (i.e., cognitive interviews). On the basis of these results, future scale designers should carefully investigate how these factors interact to reduce the likelihood of bias in surveys for children. In the empirical studies, we derived some design recommendations for pictorial scale for young school-aged children and provided ideas for further scale evaluation.

1.3. Taking the experts' perspective into consideration to design user-friendly e-assessments for children

The third objective of this dissertation was to include the experts' perspective on how to design fair and user-friendly e-assessments. To this end, we developed a series of usability/UX heuristics for e-assessment platforms targeted toward children according to the QRR approach (Quiñones et al., 2018) (named after the first authors, Quiñones, Rusu and Rusu). We derived the heuristics from an evidence-supported corpus of 506 guidelines (Gelderblom, 2008; Liu et al., 2022) and 11 heuristics (Sim, 2009). We then analyzed them thematically to synthesize them (Guest et al., 2011). We conducted a first validation study of the final 10 heuristics through two different rounds of expert reviews: a heuristic evaluation workshop and an online survey that included experts from distinct fields. The domain experts rated the heuristics on utility, relevance, clarity, and simplicity. On the basis of the outcomes, the current heuristics make a first step towards the development of a tool for evaluating e-assessment platforms designed for children between the ages of 6 and 12. Future research should continue to validate the heuristics with different e-assessment systems and in different settings to safeguard their transferability. Moreover, we advise to introduce the heuristics to more experts who can apply them to various e-assessment systems over a longer period of time. Consequently, there is a need for future research to thoroughly test the value of these heuristics. The final heuristics can then be used to improve children's overall UX and to ensure that future e-assessment platforms will be intentionally designed to meet the specific requirements and capabilities of children. In addition, we offered several recommendations for how to improve the employed approach (Quiñones et al., 2018) with a particular emphasis on enhancing the comprehensiveness of the exploration criteria from the initial phase of the approach. Therefore, in addition to emphasizing the domain and usability/UX, we propose that both the user group and technology be incorporated. Subsequent investigations should validate and augment this concept by implementing and improving the existing technique. On the basis of the considerations described above, we extended the comprehensiveness of the QRR approach by improving its ability to develop usability heuristics.

2. Contributions

This dissertation makes several contributions (Wobbrock & Kientz, 2016).

As demonstrated in all three research papers, the main contribution of this dissertation is **methodological.**

First, in our systematic review in Chapter 2 (Lehnert et al., 2022), we offer directions for future research endeavors that will facilitate the development of improved methodologies in the discipline of CCI. A systematic review offers a unique opportunity to examine the body of literature to address various concerns, and our comprehensive systematic review provides designers and researchers from different fields with a unified and cohesive synopsis of interaction design methods that are appropriate for children. Therefore, Chapter 2 makes a methodological contribution to the existing research in the field by (a) providing a comprehensive analysis of a diverse child population, including children with varying developmental abilities and those between the ages of 0 and 18 years, (b) employing various methodologies to demonstrate the interaction between the roles of children and adults, and (c) identifying patterns of triangulation in the application of research methods while considering recent ethical discussions about the participation of children in design processes.

To make our second **methodological contribution**, we evaluated child-appropriate survey methods in Chapter 3 that will contribute to a better interpretation of UX methods in the context of e-assessments. Given that extreme responses (Read et al., 2002; Zaman et al., 2012), lack of reaction, random responses, limited consistency, and low test-retest reliability in young school-aged children have previously been documented (Borgers, 2003; Borgers et al., 2004; Borgers & Hox, 2001), and confirmed in our large-scale digital questionnaire, we investigated the need for more methodological expertise in designing response scales that are appropriate for children's cognitive capacities and motivation. To address this need, in Chapter 3, we examined the thought processes that underlies children's survey answers to identify possible explanations for errors in design scales intended for children. Using a cognitive interview study, we gained insights into the complex relationships between cognitive processes, children's motivation, and the level of complexity inherent in the scale. To provide future guidance and inspiration to researchers in developing scales, we formulated six recommendations for the future development of scales intended for use with children.

Third, Chapter's 4 **methodological contribution** includes the extraction of heuristics for the design of fair and child-friendly e-assessment systems. Previous investigations of the practice of creating domain-specific sets of heuristics had prompted research investigations into the functioning of e-assessment platforms, specifically those of Computer-Assisted Assessments (CAAs) (Sim, 2009; Sim et al., 2007, 2008; Sim & Read, 2016). Nevertheless, such previous studies had overlooked the unique demographic of children and the age-related characteristics of the e-assessment application. In Chapter 4, we aimed to fill this research gap and make a valuable contribution by taking the first steps toward developing domain-specific usability heuristics for children. We achieved this contribution by (a) drawing from prior research and expert evaluation to synthesize a collection of heuristics that can be used to conduct a heuristic evaluation and (b) assessing the heuristics via two rounds of validation, including an expert survey and a workshop dedicated to heuristic evaluation. Additionally, we made several recommendations with respect to an existing framework for developing usability heuristics from which future practitioners can profit, classified as an **applied contribution** in this dissertation. Additionally, the future heuristic can guide practitioners and designers from the e-assessment community toward the design of user-friendly interfaces targeted toward children between the ages of 6 and 12. These practical contributions are important for any practitioners in the field, but we developed them specifically for our partners, DEPP and LUCET, who can profit from these heuristics and the included checklists when evaluating their e-assessment applications.

The insights from the cognitive interviews and surveys with children aimed at evaluating different response scales also led to practical recommendations for practitioners from the CCI or related disciplines (Chapter 3) and form an **applied contribution**, which offers an advantage over existing work, as we evaluated the scales in context (i.e., field study) rather than in an experimental setting.

3. Limitations

In every chapter in this dissertation, we discussed the corresponding limitations. Therefore, in this section, we direct our attention toward the overarching limitations that apply to all our studies.

3.1. Tradeoff between ecological validity and experimental biases

Introducing a research tool (i.e., scale design) in the real world (i.e., field study) versus in a laboratory setting comes with tradeoffs in the study design. A field study refers to a systematic investigation carried out in the actual environment or a naturally occurring context (Farrell & Fessenden, 2024), thus relating back to this dissertation's methodological contributions. The research settings are similar to the scenarios found in everyday life, thus maintaining the authenticity of the environment (Farrell & Fessenden, 2024) (e.g., allowing a large-scale assessment operation to examine how the variables behave in a natural environment). The primary focus for us was observing, analyzing, and describing existing phenomena rather than on actively modifying the variable in question, a focus that was important for our applied context. We benefited from conducting field research because we could directly apply the results to real-world settings (i.e., our partners' monitoring studies), as our field research covered a broader range of scenarios and environments individuals encounter in their natural surroundings (i.e., the classroom where the assessment takes place). However, with these advantages came the inability to control for certain factors (e.g., peer influence), making it difficult to draw scientifically rigorous conclusions and generalize our findings.

To address this challenge, laboratory research can be conducted in an attempt to control for extraneous factors. However, even in a careful laboratory investigation, random variables can appear. For example, in our cognitive interviews, we observed that children reported experiencing a positive emotional state due to the prevalence of an external event (e.g., forthcoming birthday party). Moreover, previous studies involving children have shown that doing research in an artificial environment, such as a laboratory setting, poses significant challenges for the study design (Hanna et al., 1997). Researchers must carefully adapt the research environment to the child's specific needs and moods (i.e., keep parents nearby). Modifying these variables can disrupt the natural environment, placing additional ethical burdens on the participants and thereby impacting their behavior and leading to changes in the outcomes. Therefore, despite the various compromises involved in conducting field studies, we decided to administer our questionnaire in a naturalistic setting and thus had to compromise on the format. We could introduce only a certain number of item/response combinations due to time constraints with the assessment setting and the cognitive demands of the children. A highly-demand questionnaire administered just before the assessment could, in turn, affect the performance of the evaluation itself. Consequently, we decided to reduce the number of items.

Moreover, to tackle the limitations of our field study and to explain the observations made with the questionnaire (i.e., the extreme positive responses), we conducted cognitive interviews. We followed an explanatory sequential design (Creswell & Creswell, 2018), a design in which the researcher first does quantitative research (i.e., questionnaire), then looks at the results and uses qualitative research (i.e., cognitive interviews) to go into more detail about what the results mean. This design may be counterintuitive because cognitive interviews or walkthroughs are usually done first so that the researcher can address the changes before the quantitative study is conducted (i.e., exploratory sequential design, (Creswell & Creswell, 2018). However, as the scales we used had already been published, and some had been used in several case studies (Kano et al., 2010; Read, 2008), we trusted their validity. In addition, our partners' interests and the yearly administration of e-assessments guided our choice to prioritize and begin with introducing the questionnaire in the field. However, as also suggested by the recent review on pictorial scales for children (Sauer et al., 2020), future studies should prioritize the development of a new scale by employing an exploratory sequential design. This approach gives the researcher the chance to adjust the variables (i.e., scale design) and test for their validity before administering them in the field. Consequently, the prevailing inclination to use natural settings rather than laboratory conditions can be attributed to the practicality and indispensability of evaluating a product solution in a real-world context (Jensen & Skov, 2005).

3.2. Generalizability of results

Children represent a distinct demographic cohort characterized by their individual and specific attributes, for instance, their age (Piaget, 1955). Thus, it remains challenging to extrapolate findings derived from a particular target group to another group, especially while assessing attitudinal and emotional aspects. As outlined in Section 1.2 of this chapter, we expected biases in children's replies due to various factors, such as their age (Read & MacFarlane, 2006; Zaman et al., 2012). Specifically, such biases could be related to the influence of the presence of the researcher or the parents or legal guardians (e.g., during our online interviews) and its impact on the research conclusions. Research conducted on adult participants has demonstrated that an interviewer's or researcher's presence can influence a study's outcomes (Rosenthal, 1977). Children are frequently viewed as being more prone to giving biased answers in comparison with adults, even in situations when researchers do not intentionally manipulate the conditions (Ceci et al., 1987). Their responses to survey questions may therefore be subject to their desire to seek acceptance or to conform to the expectations of other individuals, such as the researcher in our interviews (Read, 2008). In this regard, satisfaction and suggestibility have been identified as concerns for the validity and reliability of children's responses to survey questions (Read & MacFarlane, 2006b). Satisficing is when a participant provides relatively superficial responses that appear to be reasonable or acceptable. Children might give these responses without fully engaging in the comprehensive process of question-answering (Tourangeau et al., 2000; Tourangeau & Rasinski, 1988). In line with this concept, in Chapter 3, we discussed two biases that could potentially disrupt children's survey responses (i.e., nonresponse or extreme responding). The highly positive answers we found in both the digital questionnaire and cognitive interview could be explained by the satisficing bias. Extreme responding happens, for example, when a respondent changes their response before they communicate it (i.e., from the actual answer to the desired response) due to influences of, for example, social desirability or situational adequacy (Tourangeau et al., 2000). Several social and psychological factors (i.e., wanting to please parents, receiving a reward) can impact the thought processes of encoding, storing, retrieving, and reporting events, referred to as suggestibility (Ceci & Bruck, 1993). As Ceci et al. (1987) stated, "At times, these cues can set the emotional tone of the interview, and they can also convey implicit or explicit threats, bribes, and rewards for the desired answer. Children are quick to pick up on the emotional tones in an interview and to act accordingly" (p. 297). The complexities surrounding the concepts of satisficing and suggestibility in children's survey research provide challenges with respect to understanding and addressing these concepts (Borgers, 2003; Borgers et al., 2004; Borgers & Hox, 2001).

Previous studies have explained the biases involved in survey research by focusing primarily on age as a decisive factor (Borgers et al., 2000; Bruck et al., 1997). For example, researchers found that younger children, ages 8–10 vs. 9–10 (Read, 2008) and 7–9 vs. 12–13 (Read & MacFarlane, 2006) scored higher on the SOS than older ones. However, (Bruck et al., 1997) also emphasized that age alone cannot explain all the differences that have been observed in suggestibility. The psychosocial characteristics of adult research participants, namely, their anxiety levels and authoritarian tendencies, have been found to provoke diverse behaviors from participants in response to the researcher (Rosenthal, 1977). Consequently, other researchers should carefully scrutinize the validity and applicability of our findings, especially because children show this satisficing behavior more frequently than adults do. Therefore, the assessment of attitudinal factors remains a complex task, thus rendering the generalization of study findings challenging.

Nevertheless, it is important to investigate these psycho-motivational variables in the area of CCI. This dissertation makes a valuable contribution by identifying research needs that emerged from a comprehensive analysis of the existing literature explicitly related to the development of pictorial scales for children. An examination of the challenges associated with Interaction Design Methods for children, specifically the utilization of the survey design as a unique approach for assessing UX in an e-assessment setting, has identified a potential avenue for enhancing comprehension and guiding future research efforts in the area of CCI. Furthermore, by standardizing and implementing digital questionnaires that do not require researchers to intervene, we helped minimize potential biases in children's responses. Last, by using expert judgments, we successfully established specific guidelines for assessing children's UX in the context of e-assessments, without directly involving the children. On the basis of our findings, in the next section, we outline our vision for future research in the field.

4. Future perspectives

In this dissertation, my coauthors and I examined the notion of UX from two different viewpoints: that of children and that of experts. I argue that, particularly in light of the existing methodological limitations that surround the use of surveys to study children's self-reported UX (Chapter 3), several potential strategies can be applied to address these limitations.

4.1. Questionnaire and scale design

On the basis of our systematic review (Chapter 2), we pointed out the need to develop a questionnaire to measure children's UX during e-assessment (Chapter 3). As a first step, we selected two commonly used pictorial scales to measure children's UX; the Smileyometer Scale (SOS) (Read & MacFarlane, 2006b), and the Thumbs-Up Scale (TUS) (Kano et al., 2010). Additionally, we selected the Nodding Heads Scale (NHS) which is used to assess school attitudes (e.g., math anxiety) in the Luxembourgish School Monitoring Programme (LUCET, 2023). The authors of the SOS and TUS studied the respective scale's psychometric features (Kano et al., 2010; Read et al., 2002; Read, 2008; Read & MacFarlane, 2006; Zaman et al., 2012), whereas no evaluation studies have been conducted on the NHS. A recent review of pictorial scales in research and practice suggested that additional processes and methodologies be applied before assessing a scale's validity and reliability (Sauer et al., 2020). According to this review, a pictorial scale should be developed in three phases: (1) item generation using brainstorming and rapid prototyping, (2) interpretation check using interviews and a think-aloud protocol, and (3) scale validation using quantitative psychometric properties (Sauer et al., 2020). To our knowledge, none of the scales (SOS, TUS, NHS) have conducted evaluation studies for understanding users' mental models. Children participated in a co-design study for the SOS, which altered the face design (JRead, 2008); this research falls under item generation according to Sauer's et al. review (2020). To my knowledge, no user studies have been conducted for any of the three scale designs that address children's mental models, thus Phase 2 according to Sauer's et al. (2020) approach. Consequently, more qualitative research is needed to determine why a scale might not work for a particular age group and to explain the previously observed biases in children's responses. My coauthors and I took the first steps in this direction by investigating and evaluating suitable pictorial scale designs for children. It is up to future researchers to continue this work and investigate whether the selected scale designs (e.g., SOS) or changes to an existing scale (e.g., as summarized in the recommendations in Chapter 3) will ultimately lead to better validity and reliability. In our cognitive interviews we have additionally seen great potential and some innovative design ideas of the children for developing new scale design. Therefore, a workshop including various brainstorming techniques and co-design techniques can bring up some new and unexpected scale design ideas for phase 1 (Sauer et al., 2020). Respectively, the forthcoming scale designs should undertake a careful interpretation check, for instance, by following a similar protocol as our cognitive interview study in Chapter 3, phase 2 (Sauer et al., 2020). After iterating on the scale designs (phase 1 and 2), a final check for their psychometric properties (i.e., validity) should be undertaken, phase 3 (Sauer et al., 2020).

Moreover, regarding the items included for the development of a questionnaire measuring children's UX during e-assessment, the concept of UX in general was found to be useful for analyzing children's subjective experiences. Particularly in research with child participants, the concepts of pleasure and satisfaction (Hassenzahl, 2001), has recently captured attention in the context of learning and performance among children and adolescents. Thus, in the CCI field, concepts such as engagement (Druin, 1999) and fun (Tisza & Markopoulos, 2023) have risen in popularity, and tools have been developed to measure the underlying concepts. Thus, the design of a scale targeted toward evaluating these concepts could aid the evaluation of e-assessment experiences. However, studies would first have to investigate how important the fun component is for increasing the e-assessment experience and whether there are other concepts that should be measured (e.g., self-efficacy). Prior studies have demonstrated that students' motivation to learn or complete the assessment task can be improved by incorporating enjoyable and engaging elements, more particular, game-like elements (e.g., journeying through a game or frequent reporting on student progress) into the assessment context (Connolly et al., 2012; McClarty et al., 2012). Hence, gameful assessments offer a novel method for measuring and cultivating twenty-first century knowledge and skills (Gee, 2007).

In this regard, serious gaming has shown the potential to produce more valid assessment outcomes in comparison with traditional assessment methods because game-like features enhance the elements of enjoyment and engagement (Kato & de Klerk, 2017). However, more studies are needed to support the assertions that serious games can be a more legitimate way to test concepts and behaviors than the current methods; in addition, these studies need to determine what characteristics of these games contribute to or even reduce their validity (Kato & de Klerk, 2017). The challenge behind this approach comes from the fact that it typically needs a control group (i.e., an e-assessment application with and without gamified elements) and well-planned study protocols (Girard et al., 2013).

Moreover, children—including special target groups, such as children with ADHD—derive significant benefits from serious games due to its engaging components that effectively maintain children's attention during tasks (Kato & de Klerk, 2017). Therefore, it is also possible that other groups of students (e.g., students who suffer from test anxiety) might also benefit in similar ways. These notions have to be kept in mind when designing for inclusivity in a classroom context.

On the contrary, it is essential to acknowledge that an excessive emphasis on fun in assessing students may overshadow educational objectives, thus undermining the validity of the tests. In our particular collaboration with partners, we had to exercise caution when including fun in the context of yearly assessment procedures. Educators stated that such enhancements could inadvertently convey the notion that learning is uninteresting. Some educational professionals have argued that assessments should retain their inherent characteristics as evaluative instruments (Guerreiro & Nordengren, 2018).

4.2. Exploiting the variety of user roles

On the basis of the biases inherent in the current set of tools that are being used to measure children's UX, we have argued for the need to include experts' perspectives. In this context, in Chapter 2 (Lehnert et al., 2022), we found that the children's ages influenced the roles adults played in the design process. Specifically, our findings suggest that the facilitator role, which involves some adult participation, is the most common across the different types of feedback. Furthermore, we observed that this role had the strongest association with the active feedback the children provided. By contrast, increased interactions with adults (e.g., when adults assumed the role of an expert) were correlated with a greater tendency for children to adopt a passive role. Methodologies that focus on minimizing the impact on children through passive observation tend to have a stronger emphasis on research. The discovery above is aligned with scholarly investigations into the diverse responsibilities that the children take on throughout the design stages, indicating that children actively participate in the ideation phase of design. However, during the evaluation phase, they are mostly regarded as end users (Landoni et al., 2016).

Therefore, future research is needed to examine the interactions of the roles of adults and children and their underlying methodologies in more detail. Overall, we see the potential for solid benefits from studies that move away from focusing on the expert mindset in the direction of a participatory mindset (i.e., more active roles for children), as methodologies have been developing in this area. Therefore, future research should provide an answer to the question of the emergence of this tendency and the absence of a research-driven approach coupled with a participatory attitude, as delineated in Sanders' (2008) map of design research. Nevertheless, it is imperative to foster a paradigm shift among future researchers, prompting them to critically reassess the functions assumed by adults during the design process and to swiftly exploit the variety of roles available to them. Additionally, researchers should aim to establish a more equitable distribution of power dynamics by exploring novel approaches in research-driven and participatory frameworks, thereby granting children a more assertive role in the decision-making process.

4.3. Method triangulation beyond conventional methodologies

In addition to including the perspective of the experts, method triangulation can effectively address the biases inherent in research conducted with children. In Chapter 2 (Lehnert et al., 2022), we outlined the growing inclination to employ multiple methods within a single study. Specifically, we found that almost three quarters of the papers included in our systematic review utilized a minimum of two methods, whereas 34% of the publications employed at least three methods. Similar results were found for the adult population, with 72% of all papers using a triangulation approach (Pettersson et al., 2018). Additionally, we found that a significant proportion of the papers employed a mixed methods approach, which combined qualitative and quantitative methods. This finding is consistent with the rationale for utilizing triangulation, which involves examining a given issue from multiple perspectives to obtain a more comprehensive understanding and enhance the validity of the findings (Börjesson et al., 2015).

This approach is often employed to gain deeper insights into the research problem, next to the enhanced comprehension of the outcomes derived from alternative applied methodologies (Pettersson et al., 2018). For instance, we used our cognitive interview study as a means to interpret the finding from the questionnaire as discussed earlier in the Limitations section (i.e., explanatory vs. exploratory sequential design).

Moreover, the emphasis on triangulating methods appears to be centered on the traditional approach, with one side of the spectrum consisting of research-led methods and the other consisting of design-led approaches. The potential cause of this phenomenon could be attributed to limitations in available resources or the tendency for design processes to be presented sequentially, progressing from a chosen design approach to the eventual design outcome. Therefore, future research should invest in studying alternative combinations beyond conventional methodologies in children's research. To do so, researchers could explore novel triangulation patterns that emerge from the intersection of two dimensions: design-led versus research-led approaches and varying degrees of user involvement, ranging from passive to active (i.e., described earlier in section 4.2). The integration of research-led and design-led methodologies has the potential to address certain limitations, such as confirmation bias and early commitment (Pettersson et al., 2018).

Moreover, any endeavor to extrapolate generalizable insights would be enhanced by a concurrent investigation guided by research. Examining the same issue from multiple viewpoints can expand comprehension and offer additional stimuli for advancing technologies and products.

Virtual reality (VR) and augmented reality (AR) have garnered heightened attention in recent years as emerging methods for studying user perception (Pettersson et al., 2018), and interest in these methods has surged in the education sector. Nevertheless, the use of these emerging technologies to evaluate UX has yet to be explored, presenting a potential avenue for future application in the domain of e-assessment.

Consequently, investigating the concept of UX by exploring novel combinations of methods and stakeholder engagement is a fruitful avenue for future perspectives in studying children's UX with e-assessments.

4.4. A note on ethics in research with children

Along with the participatory mindset (i.e., developing more methods that actively engage children in the design and research process) comes the practical challenges associated with ethical and administrative obligations in conducting research with children. For instance, it is imperative that informed consent forms that are given to children use clear language that the children can understand because children have similar rights about deciding to participate, understanding that they can withdraw at any time, and the option to be briefed afterward.

Working with vulnerable groups, such as children, has always created ethical issues. The systematic review in Chapter 2 (Lehnert et al., 2022) found that over half of publications did not offer informed consent, and 76.5% did not indicate ethical approval. A recent thorough examination of the top four 2018 and 2019 HCI conferences, including CHI (Conference on Human Factors in Computing Systems), supported this finding (Van Mechelen et al., 2020). Only 21% of the articles included in our research and 24.6% of CHI articles cited an Institutional Review Board or ethics committee review (Pater et al., 2021). However, it is crucial to follow ethical guidelines, and researchers must answer fundamental questions about the risks and costs to children who agree to participate in a study. Following the EU's General Data Protection Regulation (GDPR) in 2018, the Interaction Design and Children Conference (IDC) required all research papers to include a special note on ethics and children's participation (Calder et al., 2017). In this note, the papers' authors were asked to discuss the selection of child participants, consent, treatment, data sharing, and ethics. According to ACM, an ideal informed consent form is a proactive user agreement on data collection, content, and use (Gotterbarn et al., 2017).

Moreover, Van Mechelen et al. (2014) designed CHECk, a value checklist to be used before and during design activities. This tool enables CCI researchers to critically review their values when including children in design projects and identify how to communicate participatory design to children for informed consent. The discoveries are great, but CCI researchers should stress ethics while working with children. The development of an effective code of ethics should reflect community values to help researchers make ethical decisions and direct new research approaches for children.

5. General Conclusion

In this doctoral dissertation, I aimed to examine and enhance appropriate methods to evaluate school-aged children's UX in an e-assessment context, with a specific emphasis on children aged 6 to 12 years. It is paramount to enhance the understanding of children's UX, as learners possess limited cognitive resources, and substantial variations exist across different age groups. The interface's usability, characterized by its ease of comprehension, ease of use, and ease of learning, can guarantee a positive UX for children. Good usability, in turn, allows them to focus on the primary objective of the assessment.

The preliminary investigation of UX evaluation methodologies has been acknowledged as an initial first step toward attaining this goal.

Consequently, for this dissertation, my coauthors and I conducted a systematic literature review of the current tools and methodologies, specifically focusing on evaluating suitable UX techniques to meet the abovementioned objectives. At the outset, we ascertained that there was a need for more research in the domain of CCI, specifically in younger age groups, and its relevance in the framework of e-assessment requirements. Furthermore, in alignment with prior research, our investigation revealed that the current approaches employed to assess children's UX were lacking validity. Specifically, we identified nonresponse and excessive response biases as factors that impede the reliability of children's survey replies. Our position supports improving and further developing scales that are both enjoyable and easy to use for children. To overcome the limitations we identified in our research studies, we discussed potential approaches: (1) redesigning and validating existing scales to evaluate UX (e.g., engagement or fun), (2) including different stakeholder roles for adults and children to foster a paradigm shift (e.g., by exploring novel approaches in research-driven and participatory frameworks), (3) employing method triangulation to mitigate inherent biases in studies involving children (e.g., combining design-led and research-led approaches along varying degrees of user involvement). We took a first step in this direction by applying the expert perspective to design fair and user-friendly e-assessments, resulting in 10 UX/Usability heuristics that can be applied to evaluate e-assessment platforms. However, we recommend that future research endeavors thoroughly validate the heuristics and continue to develop them by advocating for participatory design trends (i.e., active user involvement).

6. References Chapter 5

Borgers, N. (2003). Response Quality in Survey Research with Children and Adolescents: The Effect of Labeled Response Options and Vague Quantifiers. *International Journal of Public Opinion Research*, *15*(1), 83–94. https://doi.org/10.1093/ijpor/15.1.83

Borgers, N., de Leeuw, E., & Hox, J. (2000). Children as Respondents in Survey Research: Cognitive Development and Response Quality 1. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, *66*(1), 60–75. https://doi.org/10.1177/075910630006600106

Borgers, N., & Hox, J. (2001). Item Nonresponse in Questionnaires with children. *Journal of Official Statistics*, *17*(2), 321–335.

Borgers, N., Hox, J., & Sikkel, D. (2004). Response Effects in Surveys on Children and Adolescents: The Effect of Number of Response Options, Negative Wording, and Neutral Mid-Point. *Quality & Quantity*, *38*(1), 17–33. https://doi.org/10.1023/B:QUQU.0000013236.29205.a6

Börjesson, P., Barendregt, W., Eriksson, E., & Torgersson, O. (2015). Designing technology for and with developmentally diverse children: A systematic literature review. *Proceedings of the 14th International Conference on Interaction Design and Children - IDC '15*, 79–88. https://doi.org/10.1145/2771839.2771848

Bruck, M., Ceci, S. J., & Melnyk, L. (1997). External and internal sources of variation in the creation of false reports in children. *Learning and Individual Differences*, *9*(4), 289–316. https://doi.org/10.1016/S1041-6080(97)90011-X

Calder, A., IT Governance (Organization), Privacy Team, European Union, & European Parliament. (2017). *EU general data protection regulation (GDPR): An implementation and compliance guide*.https://www.itgovernanceusa.com/download/EU-GDPR-Implementation-and-Compliance-Guide.pdf

Ceci, S. J., & Bruck, M. (1993). Suggestibility of the Child Witness: A Historical Review and Synthesis. *Psychological Bulletin*, *113*(3), 403–439.

Ceci, S. J., Ross, D. F., & Toglia, M. P. (1987). Suggestibility of Children's Memory: Psycholegal Implications. *Journal of Experimental Psychology*, *116*(1), 38–49.

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, *59*(2), 661–686.

Creswell, J. W., & Creswell, J. D. (2018). *Mixed methods procedures. Qualitative, quantitative, and mixed methods approaches* (Fifth edition). SAGE Publications, Inc.

Druin, A. (1999). Cooperative inquiry: Developing new technologies for children with children. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems the CHI Is the Limit - CHI '99*, 592–599. https://doi.org/10.1145/302979.303166

Farrell, S., & Fessenden, T. (2024, December 1). Field Studies [Nielsen Norman Group]. *Field Studies*. https://www.nngroup.com/articles/field-studies/

Gee, J. P. (2007). *Good video games+ good learning: Collected essays on video games, learning, and literacy*. Peter Lang.

Gelderblom, J. H. (2008). *Designing Technology for Young Children: Guidelines Grounded in a Literature Investigation on Child Development and Children's Technology.* [Doctoral thesis, Pretoria, South Africa: University of South Africa].

Girard, C., Ecalle, J., & Magnan, A. (2013). Serious games as new educational tools: How effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning*, *29*(3), 207–219.

Gotterbarn, D., Bruckman, A., Flick, C., Miller, K., & Wolf, M. J. (2017). ACM code of ethics: A guide for positive action. *Commun. ACM*, *61*(1), 121–128. https://doi.org/10.1145/3173016

Guerreiro, M., & Nordengren, C. (2018). "No fun games": Engagement effects of two gameful assessment prototypes. *Journal of Research on Technology in Education*, *50*(2), 134–148.

Guest, G., MacQueen, K. M., & Namey, E. E. (2011). *Applied thematic analysis*. sage publications.

Hanna, L., Risden, K., & Alexander, K. (1997). Guidelines for usability testing with children. *Interactions*, *4*(5), 9–14.

Hassenzahl, M. (2001). The effect of perceived hedonic quality on product appealingness. *International Journal of Human-Computer Interaction*, *13*(4), 481–499.

Jensen, J. J., & Skov, M. B. (2005). A review of research methods in children's technology design. *Proceeding of the 2005 Conference on Interaction Design and Children - IDC '05*, 80–87. https://doi.org/10.1145/1109540.1109551

Kano, A., Horton, M., & Read, J. C. (2010). Thumbs-up scale and frequency of use scale for use in self reporting of children's computer experience. *Proceedings of the 6th Nordic Conference on Human-Computer Interaction Extending Boundaries - NordiCHI '10*, 699. https://doi.org/10.1145/1868914.1869008

Kato, P. M., & de Klerk, S. (2017). Serious games for assessment: Welcome to the jungle. *Journal of Applied Testing Technology*, *18*(S1), 1–6.

Landoni, M., Rubegni, E., Nicol, E., & Read, J. (2016). How Many Roles Can Children Play? *Proceedings of the 15th International Conference on Interaction Design and Children - IDC ’16*, 720–725. https://doi.org/10.1145/2930674.2932222

Lehnert, F. K., Niess, J., Lallemand, C., Markopoulos, P., Fischbach, A., & Koenig, V. (2022). Child–Computer Interaction: From a systematic review towards an integrated understanding of interaction design methods for children. *International Journal of Child-Computer Interaction*, *32*, 100398. https://doi.org/10.1016/j.ijcci.2021.100398

Liu, F., Sherwin, K., & Budiu, R. (2022). *UX Design for Children (Ages 3-12)* (4th ed.). CA: Nielsen and Norman Group.

LUCET. (2023). Retrieved 8 February, 2024. *Épreuves Standardisées*. http://www.epstan.lu

McClarty, K. L., Orr, A., Frey, P. M., Dolan, R. P., Vassileva, V., & McVay, A. (2012). A literature review of gaming in education. *Gaming in Education*, *1*(1), 1–35.

Pater, J., Coupe, A., Pfafman, R., Phelan, C., Toscos, T., & Jacobs, M. (2021). Standardizing Reporting of Participant Compensation in HCI: A Systematic Literature Review and Recommendations for the Field. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–16. https://doi.org/10.1145/3411764.3445734

Pettersson, I., Lachner, F., Frison, A.-K., Riener, A., & Butz, A. (2018). A Bermuda Triangle?: A Review of Method Application and Triangulation in User Experience Evaluation. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI ’18*, 1–16. https://doi.org/10.1145/3173574.3174035

Piaget, J. (1955). *The Construction of Reality in the Child*. NYBasic Books.

Quiñones, D., Rusu, C., & Rusu, V. (2018). A methodology to develop usability/user experience heuristics. *Computer Standards & Interfaces*, *59*, 109–129.

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. (2006a). 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. C., & MacFarlane, S. (2006b). Using the fun toolkit and other survey methods to gather opinions in child computer interaction. In *Proceedings 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. (2002). Endurability, Engagement and Expectations: Measuring Children's Fun. *(Eindhoven: Shaker Publishing. 2002)*, 189–198.

Rosenthal, R. (1977). Biasing effects of experimenters. *Institute of General Semantics*, *34*(3), 253–264.

Sanders, L. (2008). ON MODELINGAn evolving map of design practice and design research. *Interactions*, *15*(6), 13–17. https://doi.org/10.1145/1409040.1409043

Sauer, J., Baumgartner, J., Frei, N., & Sonderegger, A. (2020). Pictorial scales in research and practice. *European Psychologist*.

Sim, G. (2009). *Evidence based design of heuristics: Usability and computer assisted assessment*. [Doctoral thesis: University of Central Lancashire].

Sim, G., & Read, J. C. (2016). Using computer-assisted assessment heuristics for usability evaluations. *British Journal of Educational Technology*, *47*(4), 694–709.

Sim, G., Read, J. C., & Holifield, P. (2008). *Heuristics for evaluating the usability of CAA applications*. IN: Khandia, F. (ed.). 12th CAA International Computer Assisted Assessment Conference : Proceedings of the Conference on 8th and 9th July 2008 at Loughborough University. Loughborough : Loughborough University, 283-294.

Sim, G., Read, J. C., Holifield, P., & Brown, M. (2007). *Heuristic Evaluations of Computer Assisted Assessment Environments*. In *EdMedia+ Innovate Learning*, 896-905. Association for the Advancement of Computing in Education (AACE).

Spiel, K., Frauenberger, C., Keyes, O., & Fitzpatrick, G. (2019). Agency of Autistic Children in Technology Research—A Critical Literature Review. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *26*(6), 1-40. https://doi.org/10.1145/3344919

Spiel, K., & Gerling, K. (2021). The Purpose of Play: How HCI Games Research Fails Neurodivergent Populations. *ACM Transactions on Computer-Human Interaction (TOCHI), 28*(2), 1–40. https://doi.org/10.1145/3432245

Tisza, G., & Markopoulos, P. (2023). FunQ: Measuring the fun experience of a learning activity with adolescents. *Current Psychology*, *42*(3), 1936–1956.

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive Processes Underlying Context Effects in Attitude Measurement. *Psychological Bulletin*, *103*(3), 299–314.

Tourangeau, R., Rips, J. L., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge University Press.

Van Mechelen, M., Baykal, G. E., Dindler, C., & Eriksson, E. (2020). 18 Years of ethics in child-computer interaction research: A systematic literature review. *Proceedings of the Interaction Design and*

Children Conference, 161–183.

Van Mechelen, M., Sim, G., Zaman, B., Gregory, P., Slegers, K., & Horton, M. (2014). Applying the CHECk tool to participatory design sessions with children. *Proceedings of the 2014 Conference on Interaction Design and Children - IDC '14*, 253–256. https://doi.org/10.1145/2593968.2610465

Wobbrock, J. O., & Kientz, J. A. (2016). Research contributions in human-computer interaction. *Interactions*, *23*(3), 38–44.

Zaman, B., Abeele, V. V., & Grooff, D. D. (2012). Measuring product liking in preschool children: An evaluation of the Smileyometer and This or That method. *International Journal of Child-Computer Interaction*, *1*(2), 61–70.