Guidelines for Remote Usability Testing of Children's Interactive Products

A dissertation presented to

the faculty of

The Patton College of Education of Ohio University

In partial fulfillment

of the requirements for the degree

Doctor of Philosophy

Amirah M. Alkhawajah

August 2018

© 2018 Amirah M. Alkhawajah. All Rights Reserved.
This dissertation titled
Guidelines for Remote Usability Testing of Children's Interactive Products

by

AMIRAH M. ALKHAWAJAH

has been approved for
the Department of Educational Studies
and The Patton College of Education by

David R. Moore
Professor of Educational Studies

Renée A. Middleton
Dean, The Patton College of Education
Abstract

ALKHAWAJAH, AMIRAH M., Ph.D., August 2018, Instructional Technology

Guidelines for Remote Usability Testing of Children's Interactive Products

Director of Dissertation: David R. Moore

Instructional designers are often discouraged from testing their learning products due to the challenges faced when attempting to apply the traditional approaches of usability testing. As a result, scholars have been trying new methods to overcome those challenges. Remote usability testing, often operated by crowdsourcing websites, is one of those methods. However, crowdsourced usability testing websites are usually restricted to adult users. Consequently, the instructional designers of children’s learning products do not have the advantage of harnessing the potentials of remote usability testing.

As a first step for the instructional designers of children’s products to overcome the challenges of the existing usability testing methods, this study was intended to create a set of guidelines for the design of remote usability testing websites for children’s interactive products. Following a design and development research methodology, the researcher has used the literature’s guidelines of usability testing with children and the principles of cognitive load theory in designing a remote usability testing tool to answer the following research questions: What can be learned about the remote usability testing guidelines in terms of the features and characteristics of the usability testing tool that are related to the children nine to ten years old? What can be learned from iterative design process to inform remote usability testing guidelines in terms of improving the role of parents in such a way that can reduce their workload while maintaining a good quality of
usability testing results? And what concerns do parents have regarding their children’s participation in crowdsourced usability testing?

Observations, interviews, and fieldnotes were used to collect data in three iterations from 18 participants, nine parents and nine children. The results cover the specifications in the remote usability testing tool that worked with the children and the parents, and it covers the guidelines that are derived from the parents’ concerns regarding their children’s participation in remote usability testing. The results were summarized in a list of tentative guidelines for practitioners who are interested in designing crowdsourced or remote usability testing websites for children’s interactive products.
Acknowledgments

I would like to express my sincere gratitude to my advisor Dr. David Moore for his inspiration, guidance, and encouragement. Without his persistent support, this dissertation would not have been possible. I would also like to express my special gratitude to my committee members, Dr. Candice Thomas-Maddox, Dr. Gordon Brooks, and Dr. Greg Kessler, for their contributions, time, and efforts through serving on this committee.

My acknowledgment would be incomplete without thanking the people who truly made it possible. Special thanks to my wonderful husband and daughters for their endless support, patience, and sacrifice. I cannot find words to thank my parents and siblings for all the support and encouragement, and for showing faith in me. Last, but not least, special thanks to my friends Fatima Albrahim, Heyam Abo Alasrar, Nasiha Arif, and Rawya Aljarid for making this journey much more pleasant.
Table of Contents

Abstract ......................................................................................................................... 3
Acknowledgments ........................................................................................................... 5
List of Tables .................................................................................................................. 12
List of Figures ................................................................................................................. 13
Chapter 1: Introduction ................................................................................................. 14
  Background of the Study .............................................................................................. 14
  Statement of the Problem ............................................................................................. 17
  Research Questions ..................................................................................................... 19
  Significance of the Study ............................................................................................. 20
  Limitations and Delimitations of the Study ................................................................. 21
Definition of Terms ......................................................................................................... 22
  Children ....................................................................................................................... 22
  Crowdsourcing ........................................................................................................... 23
  Design and development research .............................................................................. 23
  Formative evaluation .................................................................................................... 23
  Instructional systems design ....................................................................................... 24
  Product ........................................................................................................................ 24
  Safety .......................................................................................................................... 24
  Usability ...................................................................................................................... 24
  Usability testing ......................................................................................................... 24
Summary ......................................................................................................................... 25
Chapter 2: Literature Review

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>26</td>
</tr>
<tr>
<td>Cognitive Load Theory</td>
<td>26</td>
</tr>
<tr>
<td>Cognitive structure</td>
<td>28</td>
</tr>
<tr>
<td>Cognitive load</td>
<td>29</td>
</tr>
<tr>
<td>Implications on instructional design</td>
<td>30</td>
</tr>
<tr>
<td>Worked example effect</td>
<td>31</td>
</tr>
<tr>
<td>Split-attention effect</td>
<td>32</td>
</tr>
<tr>
<td>Modality effect</td>
<td>32</td>
</tr>
<tr>
<td>Redundancy effect</td>
<td>33</td>
</tr>
<tr>
<td>Expertise reversal effect</td>
<td>33</td>
</tr>
<tr>
<td>Guidance fading effect</td>
<td>34</td>
</tr>
<tr>
<td>Imagination effect</td>
<td>34</td>
</tr>
<tr>
<td>Element interactivity effect</td>
<td>34</td>
</tr>
<tr>
<td>Isolated interacting elements effect</td>
<td>35</td>
</tr>
<tr>
<td>Relationship to usability</td>
<td>35</td>
</tr>
<tr>
<td>Usability Testing</td>
<td>37</td>
</tr>
<tr>
<td>Origins and history of usability testing</td>
<td>38</td>
</tr>
<tr>
<td>Usability</td>
<td>42</td>
</tr>
<tr>
<td>The role of usability testing in instructional design</td>
<td>44</td>
</tr>
<tr>
<td>The need for new approaches</td>
<td>45</td>
</tr>
<tr>
<td>New approaches to usability testing</td>
<td>48</td>
</tr>
<tr>
<td>Crowdsourced usability testing</td>
<td>53</td>
</tr>
</tbody>
</table>
Usability Testing with Children................................................................. 63
Importance of usability testing for and with children................................. 63
Challenges of usability testing with children.............................................. 65
Methods of usability testing with children............................................... 69
Recording and logging methods............................................................ 69
Observation methods.............................................................................. 70
Verbalization methods............................................................................ 70
The Wizard of Oz method......................................................................... 72
Survey methods....................................................................................... 73
Diary method........................................................................................... 74
Inspection methods.................................................................................. 75
Lab versus field in usability testing with children..................................... 76
Adults’ influence...................................................................................... 78
Guidelines of Usability Testing with Children .......................................... 79
Ethical Considerations............................................................................. 89
Summary.................................................................................................. 91
Chapter 3: Methodology ......................................................................... 93
Design and Development Research Methodology...................................... 93
Design and Development Research Study Design..................................... 95
Analysis................................................................................................... 95
Design..................................................................................................... 97
Development and pilot study................................................................. 100
Evaluation............................................................................................... 104
Participants ................................................................................................................. 106

Data Collection and Analysis ......................................................................................... 111
  Data collection procedure. ............................................................................................... 113
  Data analysis. ................................................................................................................... 117

The Role of the Researcher ............................................................................................ 118

Methodological Issues .................................................................................................... 119

Ethical Consideration ..................................................................................................... 122

Summary ........................................................................................................................ 123

Chapter 4: Findings ......................................................................................................... 124
  Introduction ..................................................................................................................... 124

Children-Related Features ......................................................................................... 125
  What happened? .............................................................................................................. 125
  What went well and what didn’t? ................................................................................... 126
    Introduction stage .......................................................................................................... 127
    During-the-test stage .................................................................................................... 131
    Finishing-up-stage ....................................................................................................... 152

Parents-Related Features ............................................................................................ 153
  What happened? .............................................................................................................. 153
  What went well and what didn’t? ................................................................................... 156
    Introduction stage .......................................................................................................... 156
    Set-up and planning stage ............................................................................................ 157
    During-the-test stage .................................................................................................... 165
    Finishing-up-stage ....................................................................................................... 184
List of Tables

Table 1: Hanna’s Guidelines of Usability Testing with Children ........................................ 81
Table 2: Designing the Crowdsourced Usability Testing Tool Based on the CLT .......... 98
Table 3: Participant's Details .......................................................................................... 109
Table 4: Children's Answers in First Iteration ............................................................... 138
Table 5: Children's Answers in Second Iteration ........................................................... 142
Table 6: Children's Answers in Third Iteration ............................................................... 147
Table 7: Parent's Reports in the First Iteration ............................................................... 169
Table 8: Parent's Reports in the Second Iteration ........................................................... 173
Table 9: Parent's Reports in the Third Iteration ............................................................... 176
Table 10: Time Spent Testing the Product .................................................................... 194
Table 11: Number of Reported Usability Problems by Parents in Each Iteration
........................................................................................................................................... 203
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>The general structure of children’s crowdsourced usability testing</td>
<td>102</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Data collection procedure</td>
<td>116</td>
</tr>
<tr>
<td>Figure 3</td>
<td>A problem with the timing of the visual and the auditory information</td>
<td>164</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

Background of the Study

Instructional design is an inherently systematic process utilizing feedback and iteration for primary design decisions. Usability testing is a fundamental procedure conducted by instructional designers and individuals or organizations interested in product development. Through usability testing, instructional designers can improve the usability of a learning product by detecting the problems encountered by typical users (Barnum, 2011; Rubin & Chisnell, 2008). The results of a well-conducted usability test lead to a product that “helps users achieve their goals in the simplest way possible,” or in other words, the results lead to a usable product (Nielsen & Tahir, 2000, p. 43). On the other hand, when usability testing is inappropriately conducted, intended audiences are exposed to low-quality products that might have critical problems that prevent them from being beneficial or even usable.

Scholars are concerned about the challenges faced when attempting to apply typical/lab usability testing, which often discourage some product designers from conducting the tests (Harty, 2011; Martin, Al Shamari, Seliaman, & Mayhew, 2014; Moore, 2008; Nielsen, 1993a). The main challenges of typical usability testing can be summarized in three points: it is difficult to recruit testers who have the same characteristics as the typical users of a product; the process of typical usability testing is expensive; and the process is time-consuming (El-Halees, 2014; Harty, 2011; Moore, 2008; Rubin & Chisnell, 2008; Rukshan & Baravalle, 2012). While Nielsen (1993a) assures that usability testing is indeed worth the cost, some of these challenges led designers to refrain from conducting the tests. This fact led some scholars and
practitioners to investigate the potentials of new usability testing approaches (El-Halees, 2014; Harty, 2011; Ma, Yan, Chen, Zhang, Huang, Drury, & Wang, 2013; Rukshan & Baravalle, 2012; Symonds, 2008; Trivedi & Khanum, 2012; Williams Van Rooij, 2013; Zogaj, Bretschneider, & Leimeister, 2014). Some of the latest investigated usability testing methods are based on online approaches, such as crowdsourcing (Liu, Bias, Lease, & Kuipers, 2012; Sherief, Jiang, Hosseini, Phalp, & Ali, 2014).

Generally speaking, online usability testing allows the instructional designer to recruit testers and manage the usability test through the internet. It is deemed promising in terms of overcoming some problems of traditional usability testing (Andrzejczak & Liu, 2010; Barnum, 2011; Harty, 2011; Martin et al., 2014; Moore, 2008; Rubin & Chisnell, 2008; Symonds 2011; Trivedi & Khanum, 2012; Zogaj et al., 2014). It does not only cut the costs tremendously, but it also crosses the barriers of time and geographic distances and can facilitate the recruitment of more diverse demographic groups compared to the traditional approaches. Although several usability testing websites have been launched in the last decade, such as Applause.com, Usabilla.com, and Loop11.com, these websites have a major flaw for instructional designers; all of the reviewed websites in the current study require the users to be at least 18 years old to be considered eligible participants. Consequently, the designers of children’s products can only use these services to test their products with adults rather than children, or they are not able to benefit from the services of online usability testing. This point forces the serious designers of children’s products, who actually care about conducting appropriate tests, to deal with the challenges posed by the other usability testing approaches rather than taking advantage of online usability testing potentials.
Because testing the usability of children’s products has even more challenges than testing the usability of adults’ products (Khanum & Trivedi, 2012; Markopoulos, Read, MacFarlane, & Hoysniemi, 2008), it is important to exploit the benefits and explore the feasibility of online usability testing of children’s products. On the other hand, testing children’s products with adults rather than children violates the essence of usability testing (Khanum & Trivedi, 2012; Markopoulos et al., 2008; Massion & Nam, 1998). After all, the intended users are the center of usability testing, which is conducted to evaluate user-centered designs (Rubin & Chisnell, 2008; Teixeira, Santos, & Respício, 2013; Williams Van Rooij, 2013).

Children have specific characteristics and behaviors that are known to affect the process of usability testing and that require the designer’s attention when designing the tests. For example, children can be easily distracted and feel bored (Hanna, Risden, & Alexander, 1997); they have short attention span (Rounding, Tee, Wu, Guo, & Tse, 2013); they usually have trouble in thinking aloud (Barendregt, Bekker, Bouwhuis, & Baauw, 2007; Khanum & Trivedi, 2012); they can easily forget the instructions of the task/test; some of them might try to please the designer rather than providing their true opinion (Barendregt et al., 2007); some have trouble adjusting to unfamiliar settings (Markopoulos & Bekker, 2003); their logical thinking is not mature (Markopoulos & Bekker, 2003; Khanum & Trivedi, 2012); and finally, their physical and language abilities can also influence the testing results (Markopoulos & Bekker, 2003; Markopoulos et al., 2008). Borrowing Khanum and Trivendi’s (2012) words:

Children are not just young age individuals; they represent a set of individuals who have their own perception, style, preferences, likes and dislikes. When
designing technology for children their preferences should be taken into account.

To do so, usability evaluations are performed with the children as the testers of technology (p. 101).

Therefore, it is not only tricky, but also of minimal benefit, to use adult testers in performing a usability testing for children’s products (Markopoulos et al., 2008).

Given that the current online usability testing services are limited to adult participants, it can be concluded that there is a need to develop a set of guidelines for remote usability testing with children. From the list of children’s characteristics mentioned above, applying the principles of cognitive load theory could be helpful in reducing the effects of some of these characteristics in remote usability testing environments. The guidelines provided in the current study can serve as a starting point for research that investigates the effectiveness of more specific remote usability testing approaches, such as crowdsourced usability testing. This study contributes to the literature that focuses on finding improved approaches of usability testing, specifically for children’s products.

**Statement of the Problem**

Usability testing has grabbed the attention of several crowdsourcing websites developers to overcome the challenges of the traditional approaches through remote usability testing. Despite the growth in the number of these websites, none of the reviewed instances in the current study allow the designers to test their products with children, although this age group is witnessing a rapid growth in computer and Internet access (Nielsen, 2010). Subsequently, there is little evidence that crowdsourced usability testing of children’s products has been investigated with children participants. In fact,
there is a lack of the literature concerned with usability testing with children in general as well (Byerly, 2007; Markopoulos et al., 2008; Markopoulos & Bekker, 2003; Meloncon, Haynes, Varelmann, & Groh, 2010). There should be a set of guidelines for the design of crowdsourced usability testing with children to facilitate the investigation of its feasibility and benefits, taking into account the unique characteristics of children and the capacity of online environments.

Several studies have used the guidelines of usability testing with children developed by Hanna, Risden, and Alexander (1997), which influenced the work of several researchers in the field (Barendregt & Bekker, 2003; Khanum, & Trivedi, 2012; Khanum, & Trivedi, 2013a; Khanum & Trivedi, 2013b; Markopoulos & Bekker, 2002; Markopoulos & Bekker, 2003; Meloncon et al., 2010). However, these guidelines are focused on the typical usability-testing situation that is conducted in a lab setting and might be inadequate for situations like crowdsourced usability testing or remote usability testing in general (Hanna et al., 1997). Therefore, the first step to enable the crowdsourcing of usability testing with children is to develop new practical guidelines for remote usability testing with children. These guidelines should be able to balance between reducing the cognitive load of the children and reducing the workload of the parents, while retaining a sufficient quality of usability testing results that allows the instructional designer to improve his or her product.

The objective of the present study was to create a set of guidelines for the design of remote usability testing websites for children’s interactive products. The guidelines focus on obtaining sufficient usability testing results that allow the instructional designers to properly improve their products while reducing the workload of parents who monitor
the remote usability test. In general, this study aims to identify how we can best take advantage of the potential of remote usability testing of children’s interactive products by taking the right design decisions.

**Research Questions**

In order to fill the gap in the literature, the purpose of the present study is to develop a set of guidelines for remote usability testing of children’s interactive products. A prototype of a crowdsourced usability testing website, Kid Crowd, was developed to test children’s learning products based on the literature’s guidelines of usability testing with children and the principles of cognitive load theory. Observations were made for the use of the prototype by children of nine to ten years old accompanied by their parents who were interviewed following each testing session. This narrow age range was selected because children at this age have minimum differences in their attributes compared to other age groups (Lobe, Livingstone, Olafsson, & Simões, 2008, p. 15), and at the same time, they have the ability to think aloud (Markopoulos et al., 2008) and have the computer experience that is required to participate in a usability test (Hanna et al., 1997). The data from the observations, fieldnotes, and interviews were used to answer the following research questions:

1. What can be learned about the remote usability testing guidelines in terms of:
   a. the features and characteristics of the usability testing tool that are related to the children nine to ten years old?
   b. improving the role of parents in such a way that can reduce the workload of parents while maintaining a good quality of usability testing results?
2. What concerns do parents have regarding their children’s participation in crowdsourced usability testing?

Note that the more usability problems the parents can discover in a product, the better the quality of the usability testing results.

The present study was a design and development research, which is “unique to the Instructional Design and Technology (IDT) field” (Richey & Klein, 2014a, p.141). The design and development research is defined by Richey and Klein (2014a) as “the systematic study of design, development and evaluation processes with the aim of establishing an empirical basis for the creation of instructional and non-instructional products and tools and new enhanced models that govern their development” (p. 142). Research projects of this type are concerned with the long-term advancement of the instructional design field (McKenney & Reeves, 2014).

Significance of the Study

The significance of this study lies in that it is the first step for the instructional designers of children’s products to overcome the challenges they face when conducting traditional usability testing. It proposes a new and promising approach that was only available for the other instructional designers because of the age restrictions, the lack of design guidelines, and the lack of literature. Results of this research are intended to extend the knowledge base for usability testing approaches and guidelines. The developed guidelines should provide helpful information for instructional designers on the best practices when using remote or crowdsourced usability testing websites to test children’s products. Finally, to the best of our knowledge, this is the first attempt to investigate the design of crowdsourced usability testing of children’s interactive products.
Limitations and Delimitations of the Study

There is no such thing as a perfect study; there are always “fundamental constraints of time and resources,” and that is what makes it essential to acknowledge the limitations and the delimitations of one’s study (Patton, 2002, p. 225). This is especially important in design and development research, which is often context specific (Richey, Klein, & Nelson, 2004). Possible limitations and delimitations of this study include:

1. Data validity: There is usually an issue with data validity in design and development studies because the researcher in this type of studies “also serves as the designer/developer” (Richey & Klein, 2014a, p. 148; Richey & Klein, 2005). This factor could affect the objectivity in the analysis and interpretation (Richey et al., 2004). In order to eliminate this problem, special attention was “given to instrument design, data collection and triangulating multiple sources of data” (Richey & Klein, 2014a, p. 148).

2. Data collection and integrity: The data for this study was collected through field observations, fieldnotes, and interviews only, which might lead to problems in the data integrity. Data triangulation was used to solve this problem (Richey & Klein, 2005).

3. Context and generalizability: Since design and development studies are usually context specific, “the results may be applicable only in the situation studied or to others with similar characteristics, rather than being generalizable to a wider range of instructional environments” (Richey et al., 2004, p. 1114). Context and generalizability in this study include the following points:
a. Participants: Participants were limited to children of nine to ten years old, and they participated in one usability testing session only. The study did not cover the other age ranges. The findings of this study cannot be generalized to the larger population without being replicated on other contexts and validated by future research.

b. Product and instance: This study was limited to the investigation of one children’s website using one crowdsourced usability testing prototype. The findings cannot be applied to other types of interactive products without replication and validation by future research.

4. Conclusions drawn from this study may be affected by the participants’ language skills in each iteration. It was intended to balance the number of native speakers and non-native speakers of English in each iteration, but this was not possible due to the participants’ personal schedules.

5. Conclusions drawn from this study may be affected by the participants’ gender. Since the sample was selected purposefully from the researcher’s social network, which comprised mainly of women, most of the parents in the study were women. Also, there were supposed to be three boys in the sample, but they could not participate due to family issues.

Definition of Terms

Children. Children are defined as “persons who have not attained the legal age for consent to treatments or procedures involved in the research, under the applicable law of the jurisdiction in which the research will be conducted” (U.S. Department of Health & Human Services, 2011a, para.1).
Crowdsourcing. According to Estellés-Arolas & González-Ladrón-de-Guevara (2012), crowdsourcing is defined as follows:

A type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilize to their advantage what the user has brought to the venture, whose form will depend on the type of activity undertaken. (p. 197).

Design and development research. Richey and Klein (2014a) define design and development research as “the systematic study of design, development and evaluation processes with the aim of establishing an empirical basis for the creation of instructional and non-instructional products and tools and new enhanced models that govern their development” (p. 142).

Formative evaluation. Dick, Carey, and Carey (2015) define formative evaluation as “the process designers use to obtain data for revising their instruction to make it more efficient and effective” (p. 284). It is also defined as “the iterative collection and feedback of process data to support program development and improvement” (Spector, Merrill, Elen, & Bishop, 2014, p. 957).
**Instructional systems design.** ISD is defined as the “structured process for the design, development, implementation, and evaluation of learning to improve performance and to ensure the quality of instruction with primary roots in adult education” (Spector et al., 2014, p. 959).

**Product.** A product is “part of the equipment (hardware, software and materials) for which usability is to be specified or evaluated” (ISO, 1998, p. 2). The focus of this study is on children’s computer interactive products, such as educational websites.

**Safety.** Markopoulos et al. (2008) define children’s safety in the usability testing of interactive products and communication tools, such as websites and mobile devices. An interactive product is safe for children if it does not “expose children to inappropriate Web content or cause them to give personal information to strangers” (Markopoulos et al., 2008, Chapter 5, Section 3, para. 9). In cases where children are participating in a usability testing, child abuse becomes a safety issue as well, for both the child and the evaluator (Markopoulos et al., 2008).

**Usability.** The International Organization for Standardization (ISO) defines usability as the “extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” (ISO, 1998, p. 2).

**Usability testing.** Usability testing is defined as “a process that employs people as testing participants who are representative of the target audience to evaluate the degree to which a product meets specific usability criteria” (Rubin & Chisnell, 2008, p. 21).
Summary

The present study focused on the guidelines of remote usability testing of children’s interactive products. The literature shows a need for a new usability testing approach that can overcome the challenges posed by traditional methods. Crowdsourcing is a promising remote usability testing method, but, to the best of our knowledge, it has not been tried as a tool for testing children’s products. The present study is intended to answer the following questions: What can be learned about the remote usability testing guidelines in terms of the features and characteristics of the usability testing tool that are related to the children nine to ten years old? What can be learned about the remote usability testing guidelines in terms of improving the role of parents in such a way that can reduce the workload of parents while maintaining a good quality of usability testing results? And what concerns do parents have regarding their children’s participation in crowdsourced usability testing?

The findings of the present study should extend the knowledge base for usability testing approaches, and it should provide some practical guidelines for instructional designers on the best practices for crowdsourcing usability testing of children’s products.
Chapter 2: Literature Review

Introduction

In design and development studies directed toward “innovative design and development processes, it would not be unusual to find little research in the literature that is directly relevant” (Richey et al., 2004, p. 1115). However, “the researcher must still identify literature that is relevant to the foundational theory of the project, even though the link may be indirect” (Richey et al., 2004, p. 1115). The literature review in this chapter will cover the cognitive load theory as well as usability testing, including crowdsourced usability testing, usability testing with children, usability testing guidelines, and the related ethical concerns. Because a child’s behaviors that occur in a specific setting are strongly influenced by the setting itself (Khanum & Trivedi, 2012), part of the literature review covers crowdsourcing, which is the setting or the environment in which the usability testing took place in the current study.

Cognitive Load Theory

Several studies have investigated cognitive load and its relation to the usability of computer products. It is true that the cognitive load theory is basically concerned about the design of the learning aspects of a product, but research has shown that it can also “be applied beyond learning itself to the application interface and distractions” (Deegan, 2013, p. 107). Measuring cognitive load has been used to test the usability of websites (Tracy & Albers, 2006). As Whitenton (2013) explains, the usability of websites can be maximized by minimizing the cognitive load imposed by the website’s interface. The cognitive load theory was chosen in the present study to help managing the load imposed
on the child and his or her parent, who will be administrating the usability test, by following the theory’s principles in the design of the usability testing website.

As will be shown soon, usability testing with children has several challenges. One of these challenges is the difficulty a child faces when using the think-aloud method, which could be caused by the high cognitive load imposed by the method itself (Barendregt et al., 2007; Khanum & Trivedi, 2012). The parent is also expected to experience high workload during the test since he or she will be required to perform several tasks simultaneously in the method proposed in this study. This was shown to be true in similar usability testing methods in which the parent takes the role of the evaluator or the test administrator (Markopolous et al., 2008). Therefore, designing the test based on the cognitive load theory could help in minimizing the cognitive load of both the child and the parent and in facilitating the whole process.

Cognitive load theory is concerned about the amount of instructional content and how to appropriately present it in order to be easily processed and moved into the learner’s long-term memory (Sweller, 1988; Sweller, 2002; Sweller, van Merrienboer, & Paas, 1998). “The primary purpose of the theory has been to provide a framework for instructional design” (Sweller et al., 1998, p. 265). While studying problem solving processes, John Sweller developed the cognitive load theory based on George A. Miller’s (1956) findings on the limited capacity of short-term memory, now known as working memory (Sweller, 1988; Sweller et al., 1998). The theory consists of two main parts: The human cognitive structure, which is the mental processes used to handle information; and its implications on instructional design, or the effects of different design aspects on the learning process that takes place in this structure (Sweller et al., 1998). The following
sections will review the two components of the cognitive load theory as well as the types of cognitive load.

**Cognitive structure.** Based on cognitive load theory, the human cognitive structure is organized into four components (Sweller, 2002; Sweller et al., 1998). The first component is the working memory, which “can be equated with consciousness” (Sweller et al., 1998, p. 252). Working memory is the unit where limited amount of information (Miller, 1956) can be received and kept for an extremely short amount of time (Peterson & Peterson, 1959). The capacity of working memory was found to be 7±2 items of information at a time (Miller, 1956). Working memory deals with visual information and auditory information through two different streams or channels using two different processes (Penney, 1989). This is known as dual modality. The capacity of working memory can change depending on whether one or both channels are being used at a time as will be explained shortly (Sweller et al., 1998). The design of the instructional material can either overwhelm or facilitate the job of the working memory and, consequently, the learning process.

The second component of the human cognitive structure is long-term memory. Unlike working memory, the capacity of long-term memory is extremely large and it can store complex information for years (Sweller et al., 1998). One cannot explain the job of long-term memory without talking about schema, which is the third component in the cognitive structure. Schemas are the form in which knowledge is stored in the long-term memory. The size and number of schemas determine the level of expertise of a person in a specific task (Sweller, 1988). In other words, “problem solving expertise in complex areas demands the acquisition of tens of thousands of domain-specific schemas” (Sweller,
Schemas do not only store knowledge in long-term memory, but they also reduce the load on working memory (Sweller et al., 1998). Because even complex schemas with several subschemas are considered a single item or element in the working memory, processing their sub-elements does not require a large capacity in the working memory. In working memory, “although the number of elements is limited, the size, complexity, and sophistication of elements is not” (Sweller et al., 1998, p. 256).

The fourth and the last component of the cognitive structure in cognitive load theory is schema automation (Sweller, 2002; Sweller et al., 1998). The intensive practice of a task eventually helps schemas to be processed automatically by bypassing the working memory. Therefore, automation decreases the load on the working memory because the processing of schemas in this case is done unconsciously. “Without automation, a previously encountered task may be completed, but performance is likely to be slow and clumsy” (Sweller et al., 1998, p. 258). An example would be driving the car for the first few times versus driving the car after a year of practice. Cognitive structure plays an important role in cognitive load types, which is explained next.

**Cognitive load.** Before moving to the second part of the cognitive load theory, it is important to understand the types of cognitive load and their relation to the cognitive structure. Cognitive load is “a construct representing the load that performing a particular task imposes on the cognitive system” (Sweller et al., 1998, p. 266). In simple words, it is the load placed on working memory while processing information that is not learned yet (Sweller, 2008). It consists of three parts: Extraneous cognitive load, which is imposed by distractors that are unrelated to the materials being studied; intrinsic cognitive load, which is imposed by the complexity of the materials presented; and
germane cognitive load, which is imposed by the mental effort that is associated with learning the materials and building schemas in the long-term memory (Sweller, 2008).

The learner’s performance in a task is affected by the different types of cognitive load the learner is imposed to. The goal of the instructional designer is to control the cognitive load as much as possible. Unlike the other two types of cognitive load, intrinsic load cannot be controlled through the design of the materials. Rather, it can be lowered only by learning and increasing the learners’ understanding of the materials (Sweller, 2008). On the other hand, the instructional designer can allow the learner’s working memory to be imposed to higher germane load and lower extraneous load through the appropriate design of the instructional materials. It is especially important to lower the extraneous cognitive load when the intrinsic cognitive load is high, or there will not be enough space for germane cognitive load to take place. In other words, the learner will not be able to learn the materials.

Although raising the germane load raises the overall cognitive load, it does not interfere with learning because it is an indication that schemas are under construction (Sweller, 1998). Therefore, learners should be encouraged “to invest extra effort in processes that are directly relevant to learning” (Sweller et al., 1998, p. 264). This can be achieved by developing appropriate instructional designs. The next section will discuss the second part of the theory, which is the instructional techniques learned from the cognitive load structure to enhance the effectiveness of instruction based on the cognitive load theory.

**Implications on instructional design.** The knowledge of how information is handled in the cognitive structure led to the generation of several instructional
techniques, or cognitive load theory effects, that were experimentally supported in
different populations and contexts (Sweller et al., 1998). Instructional designers are
encouraged to consider these effects as guidelines when designing technology-based
instructional materials that impose high intrinsic cognitive load and “excessive extraneous cognitive load that reduces germane cognitive load” (Sweller, 2008, p. 378).
These effects are: The worked-example effect, the split-attention effect, the modality effect, the redundancy effect, the expertise reversal effect, the guidance fading effect, the imagination effect, the element interactivity effect, and the isolated interacting elements effect (Sweller, 2008).

**Worked example effect.** Worked examples reduce extraneous cognitive load by providing the learner with the solution steps of a problem. They allow the learner to concentrate on building schemas rather than struggling to solve novel problems (Sweller et al., 1998). However, since the action required from the learner is minimal in worked examples, learners are not forced to carefully examine them, which is a serious drawback. Completion problems, on the other hand, in which “a given state, a goal state, and a partial solution are provided to learners who must complete the partial solution,” can overcome the worked examples drawback (Sweller et al., 1998, p. 275). Completion problems somehow force the learners to carefully study them, and they “decrease extraneous cognitive load, facilitate the construction of schemas, and lead to better transfer performance” (Sweller et al., 1998, p. 277; van Gog, Paas, & van Merriënboer, 2006). It is also worth mentioning that not every worked example is effective. This is what will be explained in the split-attention effect below.
**Split-attention effect.** Split attention occurs in situations in which the learner has to find information from two or more separate sources in order to understand the content (Sweller, 2008). The effect of split attention on cognitive load is trivial when intrinsic cognitive load is low (Sweller et al., 1998). Researchers found that the split attention effect is not only mediated by extraneous load, by also by germane load (Cierniak, Scheiter, & Gerjets, 2009). When the material is complicated, the process of finding the information in one source and trying to search for the other in another source in order to make sense of the content raises the extraneous cognitive load and affects learning. This rule is also applied on worked examples. For instance, if a worked example constitutes of a diagram and texts that neither of them can be understood without referring to the other, it is better to design it in such a way that integrates the information in one source rather than leaving the task of “mental integration” to the learner (Sweller, 2008). It is worth mentioning that the split-attention effect does not only apply on separated visual sources, but also on separated visual and auditory sources in terms of timing; studies found that “simultaneous presentation of the narrative and visual portions of an instructional movie resulted in better recall” and problem-solving transfer (Mayer & Anderson, 1991, p. 490). This point can be better explained in the following effect below.

**Modality effect.** It was mentioned earlier that working memory deals with visual information and auditory information through two different streams or channels, which is called dual modality. The capacity of working memory can change depending on whether visual materials, auditory materials, or both are being used. The effectiveness of integrating unintelligible information into one source is equivalent to the effectiveness of
using both visual and auditory information that is only intelligible when integrated (Sweller et al., 1998). The reason is that the load in this case is reduced on the visual working memory because it is divided between the visual and the auditory working memories. This can be done, for instance, by presenting the verbal information associated with a diagram in a spoken form. In such cases, “the use of both the auditory and visual processor can expand the effective size of working memory in an instructionally favorable manner” (Sweller, 2008, p. 376).

**Redundancy effect.** “When multiple sources of information are self-contained and can be used without reference to each other,” redundancy occurs (Sweller et al., 1998, p. 283). In redundancy effect, cognitive load is increased when redundant information sources are integrated. In other words, it is recommended to isolate information that is redundant to help decrease the extraneous cognitive load that is otherwise increased by redundancy (Sweller, 2008). This rule applies for more experienced learners, but for those with less experience it might be essential to integrate the redundant information into one source.

**Expertise reversal effect.** As the level of prior knowledge and expertise increases, “previously essential information becomes redundant and so imposes an extraneous cognitive load” (Sweller, 2008, p. 377). Therefore, effective instructional procedures, such as using explanatory texts or worked examples, might become a burden for experienced learners, even if they are effective with novice learners (Kalyuga, Ayres, Chandler, & Sweller, 2003). Based on this effect, it is recommended for the decisions on the level of details provided in technology-based instructional materials to be based on the learner’s level of expertise (Sweller, 2008).
**Guidance fading effect.** This effect occurs when the guidance gradually fades as the learner’s expertise increases. If guidance is not gradually diminished, advanced learners would be disturbed by the unneeded high level of guidance (Renkl, 2012). This is explained in the cognitive load theory as follows. As the learner gains experience, the elements that are necessary to learn a specific knowledge component are integrated into schemas, which reduces the load on the working memory (Sweller, Ayres, & Kalyuga, 2011). Therefore, previously hard to comprehend information becomes easier for more experienced learners. Guidance fading in instructional materials can be presented by providing worked examples that are then replaced with completion examples and later with full examples. Notice that each kind of examples requires different level of action on the learner part. In other words, there should be an inverse relationship between the actions required from the learner and the level of guidance provided throughout the instruction, starting with low actions and high guidance and ending with high action and low guidance (Sweller, 2008).

**Imagination effect.** This effect is useful in the design of simulations. It promotes deep processing of the learning content (Leopold & Mayer, 2015). Novice learners are asked to interact with the simulation first until they develop a sufficient level of expertise that allows them to borrow the essential knowledge from their long-term memory and process it in their short-term memory while imagining the concepts under study. This procedure was found to be “the most rapid technique for transferring information to long-term memory and so increasing levels of expertise” (Sweller, 2008, p. 378).

**Element interactivity effect.** Element interactivity is the factor that determines the levels of intrinsic cognitive load in instructional materials. It is defined by the
number of interacting elements in an instructional material, which are the elements that need to be learned concurrently in order to understand the material. A high level of element interactivity means that the instructional material imposes high intrinsic cognitive load and that it is difficult to learn. “Under these circumstances, levels of extraneous cognitive load become critical, and the cognitive load effects discussed above become relevant” (Sweller, 2008, p. 379).

**Isolated interacting elements effect.** In cases in which the intrinsic cognitive load is extremely high, it is recommended to isolate the interacting elements by presenting them individually before highlighting their interaction. Otherwise, it may be impossible to learn the instructional materials. In technology-based instructional materials, this can be achieved by providing more simplified materials rather than more realistic materials. This approach may result in limited understanding, but it is better to help the learner to initially have a limited understanding of an extremely complicated material rather than not being able to understand it at all.

**Relationship to usability.** In an attempt to see whether the concepts of cognitive load theory were integrated with the concepts of human-computer interaction, a literature review was conducted with sixty-five publications that contained cognitive load in their titles or abstract (Hollender, Hofmann, Deneke, & Schmitz, 2010). The researchers found that cognitive load theory concepts have been explicitly integrated with usability and human-computer interaction research. In fact, both cognitive load theory and usability guidelines have a mutual goal. Usability guidelines in computer software design often aim at reducing the user memory load in the designed product (van Nimwegen, van Oostendorp, Burgos, & Koper, 2006).
The literature has identified cognitive load as one of several usability themes (Nielsen, Overgaard, Pedersen, Stage, & Stenild, 2006). In other words, it is considered as one of the main kinds of usability problems. Researchers have investigated reducing the cognitive load to increase the usability of user interface (Khawaja, Chen, & Marcus, 2014; Reis et al., 2012). Hollender et al. (2010) proposed a model that describes how “existing usability principles can reduce the extraneous load induced by software usage” (p. 1285). Some researchers have used measurements of cognitive load as “an additional level of usability testing besides the normal method of watching a user interact with a site” (Tracy & Albers, 2006, p. 259).

Speaking of measurement, several scholars have been interested in measuring cognitive load (Ayres, 2006; Brunken, Plass, & Leutner, 2003; Leppink, Paas, van der Vleuten, van Gog, & van Merriënboer, 2013; Paas, Tuovinen, Tabbers, & van Gerven, 2003; Paas, van Merriënboer, Adam, 1994; van Gog & Paas, 2012). Three main methods have been identified in the literature: Psychological measures, performance-based techniques, and subjective ratings. The fact that some types of cognitive load have opposite effects on learning makes it essential to measure each type separately in order to draw sound conclusions in research. Chnotz and Kürschner (2007) identified this as a limitation of cognitive load theory that “should not be considered as a negative statement” (p. 500). They affirmed that while the psychological and performance-based methods cannot be used to measure the three types of cognitive load separately, the subjective ratings that are claimed to distinguish between the three types are unreliable. Nevertheless, since the main purpose of cognitive load theory is to serve as a framework for instructional design processes, “it should be fruitful for empirical research and for
research-based practice. A framework does not require that each theoretical construct needs its own measurement procedure” (Chnotz & Kürschner, 2007, p. 500). Going back to the relationship between cognitive load theory and usability testing, Hollender et al. (2010) pointed at a potential research area of investigating “methods for the design and evaluation of educational software systems that take CLT concepts into account” (p. 1258), which is the goal of the current study.

In this section, it was shown that the cognitive load theory consists of an explanation of the human cognitive structure and its implications on the design of instructional materials. In this theory, the cognitive load imposed on the learner is the main factor that needs to be controlled, if possible. Controlling the cognitive load through the appropriate design of the materials is important in the design of usability testing with children as will be shown in the next sections.

**Usability Testing**

As long as information and technology are dominating the daily practices in every aspect of people’s lives, “usability, as a field and as a workplace trend, is only likely to grow” (Garrison, 2013, p. 186). Numerous technological projects in the field of science and technology rely on usability testing in order to produce a robust product design. In the instructional design field, the issue of untested instructional materials was not realized until the 1960s and the 1970s (Dick, Carey, & Carey, 2015). Recently, there has been an interest in finding better ways to conduct usability testing and overcome the challenges of the more traditional approaches (e.g. El-Halees, 2014; Ma, et al., 2013; Trivedi & Khanum, 2012; Williams Van Rooij, 2013; Harty, 2011; Zogaj et al., 2014; Rukshan & Baravalle, 2012; Symonds, 2008; Sherief et al., 2014; Liu et al., 2012). This part of the
chapter will explore what the recent literature says about usability testing, including its origins and history, definitions, role in the instructional design process, the need for new approaches, and the new usability testing approaches, concentrating on crowdsourced usability testing.

**Origins and history of usability testing.** The literature indicates that the roots of usability testing go back to the human factors science and ergonomics disciplines with the first application in aerospace and automotive industries (Resnik, 2011). As the concept began to be adopted in other disciplines, more technology relevant fields started to embrace it, such as information technology, assistive technology, and medical device development (Resnik, 2011). Usability testing has been traditionally conducted through a formal process derived from the experimental research design, which requires strict controlling and sampling procedures (Barnum, 2011; Rubin & Chisnell, 2008). This approach, however, has received several critiques. According to Rubin and Chisnell (2008), it is not only difficult to meet the conditions of the experimental design, but it is also unnecessary since the goal of experimental design is to test hypotheses while the goal of usability testing is to improve specific products. Hence, it is illogical to waste money and time in recruiting a large number of randomly selected participants and conduct the test under certain conditions in order to achieve a goal that can be achieved in an informal and more flexible way (Barnum, 2011; Rubin & Chisnell, 2008).

In the 1990’s, a number of studies conducted by the usability testing experts, Jakob Nielsen, who is often called the father of usability, and Tom Landauer suggested that usability testing can be conducted informally and cost-effectively using fewer participants and still give useful results (Barnum, 2011). As a result, the discount
usability testing approach has emerged (Barnum, 2011; Rubin & Chisnell, 2008). This alternative approach does not require a hypothesis to be tested or a large number of randomly selected participants; it only requires research questions and a small representative sample of typical users who are observed while using the product in a typical context of use (Rubin & Chisnell, 2008). In this approach, it is enough to recruit only three to five participants for a usability testing. Several studies back then claimed that this number of participants is sufficient to reveal 80% to 85% of usability findings (Barnum, 2011, p. 17). However, some studies showed that these conclusions cannot be generalized because the idea that five participants are enough depends on several factors, such as the participants’ level of expertise, the usability testing method used, the type of product tested, and the participants’ mean of the overall discovery rate of usability problem (Hwang & Salvendy, 2010).

Spool and Schroeder (2001), the authors of the article “Testing Web Sites: Five Users Is Nowhere Near Enough,” are among the first researchers who refuted the rule of five users. In this study, four websites were tested by conducting 49 usability tests that consisted of one task that requires the user to purchase an item using one of the websites while being observed. The researchers then recorded the total number of usability problems discovered by each user in order as well as the number of “new” problems discovered in each time based on the problems the previous users have already discovered. They found that only 65% of the total number of problems was discovered in the first five tests. More importantly, some severe usability problems, which prevented the users from completing the task, were discovered in the test number 13 and 15. These
results show that five users are not sufficient to discover 85% of usability problems, nor is it sufficient to discover the severe ones (Spool & Schroeder, 2001).

A recent study by Hwang and Salvendy (2010) was conducted by analyzing 27 usability studies that concentrated on detecting the best number of users for usability evaluation. Hwang and Salvendy (2010) found that the three to five users rule is inaccurate. The general rule they suggested based on the analysis they performed is 10±2, whether the usability evaluation is a heuristic evaluation, cognitive walkthrough, or think aloud usability testing (Hwang & Salvendy, 2010, p. 133). They also found that the overall discovery rate of usability problems could be affected by three factors: the evaluator’s level of expertise, the length of evaluation, and the report format. Therefore, if reducing the number of users is necessary, it is crucial to improve the other factors that affect the number of problems detected in order to discover the desired percentage of problems (Hwang & Salvendy, 2010).

Hwang and Salvendy’s recommendations are parallel with the findings of a more recent study by Mäntylä (2013). In this study, the researcher has investigated the effect of the crowd size on manual software usability testing’s results. Mäntylä found that the users’ expertise affects the discovered number of usability problems. The more experienced the participant is, the more usability problems he or she can find. Another finding is that when the users of the usability testing are of lower experience, the bigger the number of users, the more usability problems they can discover. In other words, increasing the number of users can compensate for the shortage in expertise (Mäntylä, 2013). From the last two studies (i.e. Hwang & Salvendy, 2010; Mäntylä, 2013), it is
obvious that the level of expertise and the number of users can be balanced to obtain a satisfactory number of usability issues.

In 2009, and after 20 years of proposing and applying his discount usability idea, Nielsen insists that discount usability “often gives better results than deluxe usability because its methods drive an emphasis on early and rapid iteration with frequent usability input” (Nielsen, 2009, para. 6). He clarifies that the proper use of the discount usability approach is based on three components: Simplified user testing, in which 5 representative users participate in a think aloud test and provide qualitative data; narrowed down prototypes, which are paper prototypes or scenarios that are used iteratively and early in the development of the product; and heuristic evaluation, in which usability experts evaluate the product based on specific usability guidelines. In other words, using the rule of five users to conduct usability testing that does not include those three components is not considered as discount usability. This might explain the disagreement between the researchers about the validity of discount usability.

Discount usability testing is based on the idea that conducting more user tests and applying more design iterations with fewer testers is better than conducting less user tests with more testers. Although it is ideal to have more testers in order to be certain of every design iteration, the lack of resources in reality forces the designer to spread the number of testers over several iterations (Markopoulos et al., 2008). The rule of five users, however, is only successful with qualitative usability studies that include the components of discount usability, but it does not apply for other situations that require quantitative data, card sorting, or eye tracking (Nielsen, 2012a). In a study that analyzed 83 qualitative usability studies, it was found that there is a “tiny” correlation between the
number of users and the number of usability findings. “Across these many projects, testing more users didn't result in appreciably more insights” (Nielsen, 2012a, para. 13). The following section explores the definition of usability.

**Usability.** The International Organization for Standardization (ISO) defined usability as the “extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” (ISO, 1998, p. 2). In some resources, effectiveness, efficiency, and satisfaction, are called usability quality factors (El-Halees, 2014), components of usability in some (Andrzejczak & Liu, 2010), and critical measures of usability in others (Barnum, 2011). This definition has been identified as the “best-known definition of usability” (p. 11) because it does not only encompass the three critical measures of usability, but it also includes the three critical elements: specific users, specific goals, and specific context of use (Barnum, 2011).

The three critical measures of usability concentrate on the accuracy and completeness of goals achievement, how this relates to the resources expended such as time and financial costs, and how satisfied are the users of the product (ISO, 1998). Together they can give a good indication of how “usable” the product is. However, when these three elements are measured, it is important to interpret the results based on the context or the environment in which the product has been used (Trivedi & Khanum, 2012), the type of participants who have used it, and the goals or tasks that have been achieved or tested (ISO, 1998).

Usability testing, on the other hand, is “a process that employs people as testing participants who are representative of the target audience to evaluate the degree to which
a product meets specific usability criteria” (Rubin & Chisnell, 2008, p. 21). It is also defined as “the activity that focuses on observing users working with a product, performing tasks that are real and meaningful to them” (Barnum, 2011, p. 13). While the first definition focuses more on the goal of usability testing and the group of people who are included in the process, the second definition concentrates more on the process itself. Combining the two definitions give a better conception of what usability testing is.

Usability testing focuses on finding the usability problems in a product. The literature does not provide a standardized definition of usability problem (Markopoulos & Bekker, 2003). In the present study, a usability problem is a situation that results from the interaction between the user and the user interface and that prevents him or her from completing a task effectively and efficiently or that leads him or her to be unsatisfied. A real usability problem is one that “predicts a problem that users will experience in their own environment, which affects their progress toward goals and their satisfaction” (Wilson, 2007, p. 46).

It is important to understand the difference between the terms usability testing and usability evaluation. First, the main goal of usability testing is to identify real problems that prevent a product from being usable in order to perform the appropriate iterations (Barnum, 2011; Rubin & Chisnell, 2008), while the main goal of usability evaluation is to build usability in a product (Rubin & Chisnell, 2008). Second, usability testing refers to techniques in which typical users are recruited to perform the test (Rubin & Chisnell, 2008; Barnum, 2011), while usability evaluation’s techniques refer to those that are performed by typical users as well as others that are performed by usability and domain experts (Trivedi & Khanum, 2012). Thus, usability evaluation can be
categorized into two broad sets of techniques: Inspection techniques, which employ
domain and usability “experts” to evaluate the product from an expert’s viewpoint, and
test techniques, in which “typical users” are recruited to test the product (Trivedi &
Khanum, 2012). Usability evaluation techniques include, but not limited to: Heuristic
evaluation, walkthroughs (Rubin & Chisnell, 2008; Hwang & Salvendy, 2010), think
aloud (Hwang & Salvendy, 2010), paper prototyping, and surveys (Rubin & Chisnell,
2008). From this perspective, usability testing can be considered as a subset of usability
evaluation; however, it can be noticed that both terms are used interchangeably in the
literature. In the present study, we are concerned with the techniques that are performed
by the typical users of a product, namely, usability testing techniques.

The role of usability testing in instructional design. There is a connection
between usability testing and user-centered design. In user-centered design, the designer
concentrates on the users’ needs through an iterative design process. In order to do so, it
is an essential step to conduct usability testing by involving representative users
(Barendregt et al., 2007; Markopoulos et al., 2008). Usability testing here is conducted to
evaluate the user-centered design to ensure that the needs of the learners or the end users
are met (Rubin & Chisnell, 2008; Teixeira, Santos, & Respício, 2013; Williams Van
Rooij, 2013).

Usability testing also has an important role in the instructional design process in
general. It can be classified as one of the “standard instructional design tasks,” and it is
represented in different instructional design models as the formative evaluation phase
evaluation as “the process designers use to obtain data for revising their instruction to
make it more efficient and effective” (p. 284). It is also defined as “the iterative collection and feedback of process data to support program development and improvement” (Spector, Merrill, Elen, & Bishop, 2014, p. 957). Dick et al. (2015) determined six types of data that need to be collected during a formative evaluation process. At least three of these types can be collected through usability testing and are related to the effectiveness, efficiency, and satisfaction of the use of the product, which are the three critical measures of usability discussed earlier in this chapter (Barnum, 2011). In other words, a considerable portion of the formative evaluation procedure can be conducted through usability testing. Therefore, it can be said that usability testing is an essential part of the formative evaluation conducted in the instructional design process.

Instructional designers did not actually begin testing their materials until the late 1970s. Since then, and at the early stages of the adoption of this procedure, designers used to test the instructional materials only after the completion of the first draft. Later, they realized through experience that it is more beneficial to start testing the materials earlier in the instructional design process (Dick et al., 2015). Although it sounds more demanding to start the usability testing even before the first draft is completed, this approach can actually cut the costs and reduce the efforts that would have been otherwise spent on fixing materials that are already implemented with problems not discovered ahead of time (Dick et al., 2015, Nielsen, 2009). The next section will explain the need for new approaches to usability testing.

**The need for new approaches.** Usability testing outcomes are influenced by different factors, resulting sometimes in misleading interpretations, which can lead to
low-quality products with critical problems. For instance, studies have shown that the prototype medium of an application (Boothe, Strawderman, & Hosea, 2013), the type of device used to perform the test (Martin et al., 2014), the type of report used to conclude the results, the length of the testing session (Hwang & Salvendy, 2010), the number of users, the level of expertise of the evaluators (Hwang & Salvendy, 2010; Mäntylä, 2013), time pressure (Mäntylä, 2013), and the physical and social contexts (Trivedi & Khanum, 2012) are all factors that can affect the usability testing outcomes. The process of conducting a usability test that provides high-quality information about the usability of a product and that can lead to the required improvements can be tricky, considering all the different factors that might influence the test. Therefore, the results of usability testing do not always prove the usability of a product (Rubin & Chisnell, 2008).

Usability testing has also other types of challenges. First, it is hard to find users who have the same characteristics as the typical users of a product. Second, it is hard to create a usability testing experience that is identical to the real experience of a product’s typical users and contexts. Rubin and Chisnell (2008), however, clarified that even though these challenges can prevent usability testers from achieving the goals of the test, it is possible to obtain true usability outcomes by cautiously implementing the usability test.

Yet, another challenge is the high costs of typical usability testing. Some of these costs are financial, such as the costs of recruiting and compensating the right participants and the costs of preparing the lab with the necessary equipment. Recruiting specialists, for example, may cost from $125 to $300 multiplied by the needed number of users (Rubin & Chisnell, 2008). Nielsen (1993a) confirms that one round of usability testing is
expensive even without counting the costs of iteration, which includes redesigning of interface, changing the prototype, and conducting a new round of usability testing. While there are other cheaper ways, such as using Craigslist, these ways have critical problems that might negatively affect the usability testing outcomes, including cheating and repeating the tests by the same participants to collect more money. The recruitment costs do not include the compensation costs, which varies depending on the required skills and professions (Rubin & Chisnell, 2008). Additionally, the costs of preparing a usability-testing lab might be unaffordable.

Other costs are not necessarily calculated financially, such as the costs interpreted in terms of time and human resources (El-Halees, 2014). For instance, the designing and implementation of a usability-testing lab can take a whole year to be completed (Garrison, 2013). The process of recruiting the right users for the process can also be lengthy (Moore, 2008). Moreover, part of the difficulty of usability testing is attributed to the length of time spent in the process of collecting the data during the test (Harty, 2011). On the other hand, usability testing cannot be done without human resources. First, usability testing requires human users to test a product because what is actually tested is not the product per se, but “its relationship to a human being with certain specific characteristics” (Rubin & Chisnell, 2008, p. 67). This point is considered one of the challenges especially that users recruitment is required multiple times during the life cycle of the product development (Moore, 2008; Williams Van Rooij, 2013). Second, usability testing might require some other human resources other than the users, such as test moderators, observers and note takers, as well as timekeepers (Rubin & Chisnell, 2008).
Overall, typical usability testing is expensive and time consuming (El-Halees, 2014; Harty, 2011; Moore, 2008; Rukshan & Baravalle, 2012), and it requires a careful implementation to attain useful outcomes (Rubin & Chisnell, 2008). Some of these challenges led several designers and organizations to refrain from conducting usability testing (Harty, 2011; Moore, 2008), which at the end can have critical effects on the quality of their products. Fortunately, recent literature has presented several attempts to innovate new approaches to usability testing to overcome some of the reviewed challenges. The new approaches of conducting usability testing are discussed in the following section.

**New approaches to usability testing.** The recent advancements in technology have allowed several alternatives to the traditional lab-based usability testing approach to emerge. These can be divided into three categories: Asynchronous remote usability testing; synchronous remote usability testing (Barnum, 2011); and self-reporting usability testing (Rubin & Chisnell, 2008). Each approach has some advantages and disadvantages as will be shown shortly.

Automated asynchronous usability testing is argued to have several advantages over the traditional, in-person usability testing. The first and most important advantage is its extended reach and range (Barnum, 2011; Harty, 2011; Rubin & Chisnell, 2008). It allows the product designer to reach a large number of diverse users. Other benefits are the ability to collect enormous amounts of quantitative data (Rubin & Chisnell, 2008) and the increased validity of the outcomes (Barnum, 2011). However, this approach is extremely expensive since it requires software that usually costs tens of thousands of dollars, or even more (Rubin & Chisnell, 2008). In addition, this approach does not
capture the “emotional information” that can be observed during a lab testing session (Rubin & Chisnell, 2008, p. 311; Barnum, 2011), neither does it allow for extended communication between the test moderator and the users (Barnum, 2011). Harty (2013) has provided a list of challenges of using automated usability testing. This list includes the difficulty of finding an automated usability testing software that is compatible with all devices or operating systems, as well as the problem that some automated usability testing software require the test moderators to have programming skills in order to customize their tests.

A number of automated usability-testing attempts have been recorded in the literature (e.g. El-Halees, 2014; Ma et al., 2013). For instance, El-Halees (2014) presented an opinion mining automated usability testing method, which automatically analyzes the users’ opinions from written texts and provides conclusions on the effectiveness, efficiency, and satisfaction elements of a product. Similarly, Ma et al. (2013) created an automated usability testing toolkit that can be downloaded on mobile devices to automatically record the users’ actions on the user interface of the application. The researchers found that the toolkit was effective in discovering the critical problems of the tested product, but because of the lack of communication and direct observation of the users, the toolkit was unable to give any information on the cause of the discovered problems.

The literature has also provided multiple examples on asynchronous usability testing that are not automated (e.g. Liu et al., 2012; Sherief et al., 2014; Williams Van Rooij, 2013; Zogaj et al., 2014). For instance, Sherief et al. (2014), Liu et al. (2012), and Zogaj (2014) investigated crowdsourced usability testing, which will be discussed later in
Similarly, Williams Van Rooij (2013) used an online research panel to conduct an asynchronous usability testing of an instructional design project over two rounds. The participants were given an online usability-testing booklet and were asked to complete some tasks online and to return the results on a specific day. The findings imply that this approach could be viable for overcoming the recruitment problems faced in other approaches.

The second category of the alternatives to the traditional approach to usability testing is the synchronous remote usability testing. This main difference between this approach and the lab-based approach is in the spatial distance (Barnum, 2011). According to Barnum (2011), this approach has several advantages. It allows the product designers to reach potential users as well as observers who cannot come to the lab, and it extends the diversity of the users. It allows the product to be tested on the users’ own devices, and it reduces the costs that would be otherwise spent in lab-based testing or automated testing. On the other hand, there are a number of disadvantages to this approach as well (Barnum, 2011). For instance, problems might arise during the setup procedure or during the test itself due to problems in the Internet or the network. Another disadvantage is the difficulty of simultaneously observing what is happening on the screen and observing the facial expressions of the user. The literature has provided varied applications of this approach, such as using 3D virtual labs and web conference tools (Trivedi & Khanum, 2012). The findings of such studies indicate that synchronous remote usability testing can be as effective as the lab-based approach in terms of the ability to detect critical problems and the amount of time spent on the test (Trivedi & Khanum, 2012).
Self-reporting usability testing, which is the third category of the alternatives to the traditional approach, was discussed by Rubin and Chisnell (2008) and investigated by Symonds (2011) and Martin et al. (2014). This approach relies on surveys and diary studies. The users answer a questionnaire that can be either taken after completing all the required tasks or divided so the users can take it while moving from one group of tasks to another. This approach can be useful to obtain an insight on the users’ impressions and satisfaction with the product, and it can even help in understanding the different factors that might influence the users’ performance (Rubin & Chisnell, 2008). However, these kinds of questionnaires need to be conducted in an unobtrusive way to avoid leading the users to alter their behavior while performing the tasks (Rubin & Chisnell, 2008). In other words, human presence is not recommended during the test. Additionally, in cases where the questionnaires are distributed through websites like Craigslist, there is a greater chance for people to cheat by submitting multiple responses in order to collect more money (Rubin & Chisnell, 2008).

Symonds (2011) investigated the advantages as well as the disadvantages of using SurveyMonkey® as an alternative usability-testing tool. After using SurveyMonkey® to test an academic library’s digital collections website, she found several advantages that can be attributed to the approach. In addition to the benefits of remote usability testing mentioned above, Symonds (2011) found that the user-friendliness and the familiarity with the survey tool allowed for easier implementation and access of the usability test. For instance, SurveyMonkey® did not require any coding skills or installation in order to be used. Furthermore, it has some features that can be helpful after the completion of the usability test, such as time tracking and data analysis and distribution. On the other hand,
Symonds (2011) discussed some drawbacks of the tool. For instance, the quality of the users’ responses is the factor that determines the quality of the usability testing outcomes in online survey tools; if the users do not provide lengthy answers in the open-ended questions, the data collected would not be very useful. There is also the problem of the users’ distributed attention when moving between the survey’s page and the website under investigation.

Martin et al. (2014) also investigated the cost benefits of asynchronous, remote usability testing using SurveyMonkey®. Only this time the researchers compared the outcomes of two groups, one that used the online survey approach and another that used a lab to conduct a traditional think aloud usability testing. The tested website was created in a way that guided the users in the online survey group throughout the testing process using real-time support. Martin et al. (2014) found that the quality of remote testing was at least equal to the quality of the traditional testing since they both took the same amount of time, and the remote testing discovered more problems than the traditional testing did. They concluded that remote testing can be used to overcome the costs challenges “if testing is based around a single attribute, such as participants’ experience” (Martin et al., 2014, p. 104). However, lab testing is better in controlling the groups of users when “multiple attributes are required for multiple comparisons, such as age, gender and cultural backgrounds” (Martin et al., 2014, p. 104).

In summary, “usability studies are typically conducted in usability labs, however, advances in communication and networking technologies afford new types of usability testing that forgo the lab” (Andrzejczak & Liu, 2010, p. 1258). For each type of usability testing, there are both advantages and disadvantages. However, it is important to know
that remote testing does not only solve the recruitment issues, but it has also been found that it can be conducted in a way that makes it provide similar quality of outcomes, if not better, compared to the lab-based testing (Andrzejczak & Liu, 2010; Martin et al., 2014; Trivedi & Khanum, 2012).

**Crowdsourced usability testing.** Although there is an increase in the number of parties interested in using crowdsourcing for product testing (Zogaj et al., 2014), crowdsourced usability testing has not received sufficient attention in the literature yet (Sherief et al., 2014; Zogaj et al., 2014). In this approach, the product designer can resort to crowdsourcing websites and intermediaries to conduct the usability testing remotely. Although it can be generally classified as asynchronous remote usability testing, some cases of crowdsourced testing can also be classified as automated or self-reporting based on the techniques used in the process. This fact reflects the high flexibility of the crowdsourced usability testing approach.

*What is crowdsourcing?* In simple terms, the word crowdsourcing today means seeking help through the Internet from a big number of people around the globe, who are usually not part of the organization, to achieve specific goals that cannot be achieved otherwise at the same level of cost-effectiveness. Estellés-Arolas and González-Ladrón-de-Guevara (2012) followed a systematic approach to offer a comprehensive definition to the concept after analyzing and extracting the mutual elements of 40 original definitions available in the literature and then validating the one they formed. They concluded that crowdsourcing is:

A type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying
knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilize to their advantage what the user has brought to the venture, whose form will depend on the type of activity undertaken. (Estellés-Arolas & González-Ladrón-de-Guevara, 2012, p. 197).

The underlying theory of the crowdsourcing phenomenon is called the wisdom of crowds (Brabham, 2008; Surowiecki, 2004). This theory suggests that under certain conditions, the collective intelligence, decisions, and solutions of a crowd are better than those coming from one person, even if this person was an expert or was smarter than everyone else in the crowd. The conditions that must be fulfilled for a crowd to be wise are: Diversity, independence, decentralization, and aggregation. In other words, the crowd has to be composed of a heterogeneous group of people with diverse backgrounds who can share their opinions independently with a limited influence from other members in the crowd. The decisions should not come from one person or authority, but it should emerge from the crowd, which consists of people with different knowledge and expertise, and thus there should be a way to aggregate these decisions (Surowiecki, 2004). All of these four conditions should be found in any well-designed crowdsourcing initiative. The next sections will explain crowdsourcing in more depth and will synthesize what the literature says about it.
Implications of crowdsourcing in usability testing. James Surowiecki, the author of *The Wisdom of Crowds: Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations* (2004), provided numerous cases of different organizations and individuals that harnessed the wisdom of crowds through crowdsourcing to solve a variety of problems. Based on the variant cases provided in his book, Surowiecki (2004) concludes that the collective intelligence of the crowd is not limited to one type of problems and that it can be used to solve even those problems of high complexity. As professionals from different fields used crowdsourcing to help completing their field-specific jobs, product designers have also employed crowdsourcing to test their products (Liu et al., 2012; Lease, Hullman, Bigham, Bernstein, Kim, Lasecki, Bakhshi, Mitra, & Miller, 2013; Sherief et al., 2014; Zogaj et al., 2014). The literature has identified some benefits of using crowdsourcing for usability testing purposes.

Studies on website and software testing have shown that crowdsourcing is beneficial in easily recruiting larger number of participants to validate the designed products in less time and at extremely lower costs compared to the traditional approaches (Liu et al., 2012; Sherief et al., 2014). In addition, “the cumulative results of crowdsourcing may well be of greater value to an organization than a smaller number of lab tests” (Liu et al., 2012, p. 8). Researchers claim that crowdsourced usability testing provides the advantages of the diversity of both human recourses and contexts, which are two important features because together they produce scenarios that are unpredictable, and thus not tested in traditional approaches (Liu et al., 2012; Sherief et al., 2014).

The fact that the testing is not conducted in a lab setting means that the products
are allowed to be tested in their real contexts, which in turn means that real problems in
the tested product can be discovered (Sherief et al., 2014). In addition, because several
devices with different features are used during the testing process, there is a bigger
chance to discover more problems based on the used device, which is not possible in the
traditional usability testing approach (e.g. Martin et al., 2014). It also enhances the
testing process since it provides a convenient way of dividing the users into groups based
on their attributes that are provided in their profiles (Sherief et al., 2014).

Regarding the social context of usability testing, it was found that the users of lab-
based testing experience higher levels of stress, although not significantly higher, when
they think that their performance is being tested, which can affect their task performance
during the usability testing (Andrzejczak & Liu, 2010). The physical context of lab-
based usability testing might also influence the outcomes of the test (Trivedi & Khanum,
2012). For instance, the users might alter their behavior when they realize they are being
observed using a camera and other recording devices. In a crowdsourcing context,
however, the users perform the usability testing under more relaxing social and physical
contexts.

Some studies from the instructional design field started to focus on crowdsourcing
as some scholars have drawn the attention to the challenges faced by instructional
designers who strive to enhance their products by conducting usability testing. Moore
(2008) has identified some of these challenges and has suggested using crowdsourcing
intermediaries to overcome these challenges, specifically the type that applies the micro-
tasks and micro-payment methods like Amazon Mechanical Turk. The challenges
identified by Moore (2008) can be summarized as follows: The current usability testing
practices are of low efficiency which results in low-quality products; the main reason is that it is difficult to find a sufficient number of usability testers who are representative of the target audience of the designed product, especially with the high costs of this process. Some studies support these conclusions. For instance, in a research and development study conducted by Sahrir, Alias, Ismail, & Osman (2012) to create a web-based game that teaches Arabic vocabulary, one of the challenges faced by the instructional designers when they attempted to design the game was “the financial implication” which was identified as “one of the restrictions to producing a high technology games” (p. 116), leave alone the costs of testing it and creating several iterations in the design. It is true that there are considerable benefits and advantages of crowdsourcing, but we should not neglect the truth that it also has several limitations and challenges.

**Limitations and challenges of crowdsourced usability testing.** The literature has identified several limitations and challenges regarding crowdsourcing, ranging from the difficulty of starting and managing such projects to the effort needed to obtain high-quality outcomes. First, managing a crowdsourcing project is a complicated task that requires a dedicated effort to properly prepare for and administer the crowdsourcing process, the crowd, and the technology (Zogaj et al., 2014). It includes several subtasks and requires a big set of decisions that need to be made to insure a smooth and risk-free process. Some of these subtasks and decisions include

- defining the requirements of the tasks (Zogaj et al., 2014);
- deciding which quality assurance methods are to be used (Sherief et al., 2014; Liu et al., 2012; Zogaj et al., 2014);
• deciding the appropriate type of incentives that can motivate the target crowd (Sherief et al., 2014; Liu et al., 2012; Zogaj et al., 2014);
• selecting the right crowd that is capable of carrying the tasks (Sherief et al., 2014);
• ensuring privacy (Sherief et al., 2014);
• selecting methods that can decrease spam and increase the quality of the outcomes (Liu et al., 2012); and
• finding appropriate ways of aggregating the crowd’s diverse contributions (Sherief et al., 2014).

This list sheds the light on the other possible challenges and limitations of crowdsourcing besides the complexity of the management process.

One of the limitations that received a vast attention in the literature is the low quality of the product or service provided through crowdsourcing (e.g. Allahbakhsh, Benatallah, Ignjatovic, Motahari-Nezhad, Bertino, & Dustdar, 2013; Ipeirotis, Provost, & Wang, 2010; Shah & Zhou, 2014; Litman, Robinson, & Rosenzweig, 2014; Sherief et al., 2014; Liu et al., 2012; Zogaj et al., 2014). Some of the identified reasons behind producing low quality outcomes are:

• Some crowd members tend to complete as many tasks as possible in the shortest amount of time in order to gain as much monetary rewards as possible, which results in poor-quality work (Liu et al., 2012).
• Some crowd members know about the disadvantage of crowdsourcing platforms that they cannot easily identify cheating, which encourages them to cheat rather than putting greater effort in the task (Ipeirotis et al., 2010).
In cases where answers to questions are required to complete the crowdsourced task and get paid, some crowd members try to answer all questions by entering random answers in order to collect the reward, even if they do not actually know the right answer. This behavior increases the error rate in the collected data (Shah & Zhou, 2014). Since crowdsourcing comprises people from different countries, it is a challenge for some of them to understand the instructions of the required task, which sometimes results in irrelevant work (Litman et al., 2014).

Although there are doubts around crowdsourced usability testing regarding the quality of its data, there is consensus that this problem can be fixed by the cautious implementation of the process (Liu et al., 2012; Sherief et al., 2014). While Liu et al. (2012) found that the quality of the lab-based usability testing was higher than the quality of the data obtained in the crowdsourced usability testing, they also found that the crowdsourced version of the testing was able to discover some critical usability issues. More importantly, it was inadequate to compare the quality of the two methods because not only they used different tasks in each study, but they also used participants who were more familiar with the tested website in the lab-based testing, while the users of the crowdsourced testing had no familiarity with the product (Liu et al., 2012).

Another limitation in crowdsourcing is the lack of communication between the crowdsourcer and the crowd. Unlike face-to-face interactions, crowdsourcing is conducted remotely. In such cases, it becomes extremely important to write the instructions of the tasks as clear as possible to minimize misunderstanding and confusion (Liu et al., 2012). Failing to do so may result in incomplete, irrelevant, or low-quality work. Having the job done remotely has another disadvantage. In a study where two
similar usability tests were conducted, one in a lab and the other through a crowdsourcing platform, the researchers noticed that when participants were asked to think aloud in the lab settings, “they sometimes mentioned minor frustrations that they did not always remember in any detail later on” (Liu et al., 2012, p. 3). The researchers were able to record these minor frustrations only in the lab settings, but it was hard to know they even existed in the crowdsourcing study. Thus, some details could be missed when using crowdsourcing for some kinds of tasks.

Other studies have identified other limitations to using crowdsourcing for usability tests and research purposes. One of these studies sheds the light on the difficulty of controlling the experimental setting, which affects the ecological validity of such studies (Kittur, Chi, & Suh, 2008). Another study emphasized the limitation that there is a group of people who cannot participate in such studies, simply because they do not have enough computer literacy to the extent that allows them to have an account in crowdsourcing websites (Liu et al., 2012). Finally, “the literature is still limited in providing engineering approaches and foundations to develop crowdsourcing platforms for” some specific tasks, such as usability testing (Sherief et al., 2014, p. 2).

**Best Practices in Crowdsourcing.** Does crowdsourcing always improve a product? The short answer is no. Due to its complexity and limitations, whether crowdsourcing improves a product or not depends on how well the crowdsourcer has prepared and taken into consideration all the possible scenarios. There are several strategies that need to be considered to overcome the challenges and the limitations of crowdsourcing. Quality assurance is a crucial issue in conducting successful crowdsourcing projects. Several studies have examined the effect of applying different
quality control methods. There seems to be consensus that when using specific control strategies, the overall quality of the obtained outcomes increases significantly (Litman et al., 2014; Goodman et al., 2013; Sarasua, Simperl, & Noy, 2012; Shah & Zhou, 2014).

One of the methods that proved to be effective is simply asking the participants to not cheat (Goodman et al., 2013). Another method requires adjusting the pay rate based on the expected hourly wage. It was found that there is a chance of obtaining higher quality information by offering a pay rate that is slightly higher than the minimum wage (Litman et al., 2014). Another study found that “fine-tuning” the wording of the questions used in the crowdsourced tasks has a positive effect on the data quality since it reduces misunderstanding and confusion (Sarasua et al., 2012). One of the widely used quality assurance approaches is eliminating spammers by adding “gold standard” questions, which can discover those who answer the questions inattentively (Shah & Zhou, 2014).

In general, quality control approaches can be categorized into two types: approaches that are applied during the design of the task and approaches that are applied during the task itself (Allahbakhsh et al., 2013).

Finding the right balance is another crucial point for successful crowdsourcing. The right balance should be achieved between the reward and the quality of the completed work (Sherief et al., 2014; Liu et al., 2012). As discussed earlier, it has been found that the best monetary reward when using micro-payment-based crowdsourcing platforms is the one that is slightly higher than the minimum wage. The right balance should be also achieved between the simplicity of the interaction methods and the accuracy of the obtained data (Sherief et al., 2014). In other words, simplifying the
methods of obtaining the data from the crowd should not be at the expense of collecting meaningful and precise data.

Finally, the answer to the question above is as follows: Even if the crowdsourcer managed to carry a successful crowdsourcing project using a wise crowd and applying cautious procedures to improve a product, it is still the crowdsourcer’s responsibility to complete the task and translate the crowd’s outcomes into a meaningful and practical solution to improve the product. Furthermore, usability testing of a product, whether performed through crowdsourcing or not, does not guarantee a product usability, and thus it does not guarantee improvement, since it cannot be considered a real experience, and since it depends on how well the test was conducted (Rubin & Chisnell, 2008). However, the chance of producing a reliable and robust product is significantly maximized when usability testing is being cautiously conducted (Rubin & Chisnell, 2008). Crowdsourcing can facilitate some aspects of usability testing while impeding others, and thus, it could provide bigger chances for product improvement when the designer makes the right decisions and understand the inevitable trade-offs.

There are several crowdsourcing intermediaries that specialize in usability testing, such as Crowdsourcedtesting.com, Applause.com (formerly Utest), Usabilla.com, and Loop11.com. Barnum (2011) described how some of these tools provide the testers with data from several types of resources. For instance, they collect the users’ responses to surveys; they collect the data from the user’s interaction with the product’s interface; and they record the voice of the users who can be asked to think aloud while using the product under investigation. Although there are a number of crowdsourcing websites that are specially created for usability testing, there is a lack of literature that investigates their
effectiveness. This can be attributed to some limitations in the websites or to the relatively high costs compared to cheaper methodologies. For instance, Symonds (2011) decided to use SurveyMonkey® instead of Loop11.com or Usabilla.com because of the difficulty of the coding required and the limitations in the number of participants and tested pages in the free version of these platforms. Another problem with the crowdsourced usability testing websites is that they only allow people over 18 years old to participate in the tests. This means that children’s products cannot be appropriately tested through these websites. The next section will explore usability testing with children and how it is different than usability testing with adults.

**Usability Testing with Children**

The literature on design and usability testing with children is scarce in comparison to the literature on usability testing with adults. In Europe, it was found that there is a lack in research concerned with children’s use of the Internet and new media, especially for those under 12 (Staksrud, Livingstone, Haddon, & Olafsson, 2009), and the situation is probably the same in the U.S (Meloncon et al., 2010). Methodological and ethical concerns have a main role in the absence of younger children in research (Plowman, 2015). It was also found that the creation and evaluation of children’s online content is one of the most neglected topics in research (Staksrud et al., 2009). The following sections explain why it is important to understand how to properly conduct usability tests for and with children, the challenges of usability testing with children, its methods, its guidelines, and the major ethical issues surrounding it.

**Importance of usability testing for and with children.** As there is a continuous increase in children’s use of online resources (Nielsen, 2010; Staksrud et al., 2009), it
becomes critical to provide the children with products that help them achieve their goals with effectiveness, efficiency and satisfaction. Unfortunately, the design of online content is “typically based purely on folklore about how kids supposedly behave” (Nielsen, 2010, para. 3). Some designers take the easiest approach of testing their designs against usability problems by using inspection methods, which rely on adult experts’ judgment or educated guess rather than involving representative children in the testing process (Markopoulos et al., 2008). It is actually common that the product designers ignore the role of children in the testing process, which may be “due to the traditional power structure of the “all-knowing” adult and the “all-learning” child” (Druin, 2002, p. 1).

It is not an easy task for a designer to produce a usable product for adult users, and it is even harder to produce a usable product for children (Barendregt et al., 2007). The problem lies in the fact that children differ from adults in several ways; “they have developing motor skills, limited reach, short attention spans, limited exposure to traditional user interfaces and social protocols,” to name but a few (Rounding et al., 2013, p.1663). Therefore, it is hard for adults to understand how children would perceive the usability of a product because it requires an “insight into children’s changing abilities, interests, and knowledge, at different ages, and the ability to emphasize with them and have an educated opinion regarding what will be appealing in terms of aesthetics, fantasy, and so forth” (Markopoulos et al., 2008, Chapter 15, Section 5, para. 3). Even the child’s own parents cannot always understand or predict their child’s judgment of the usability of a product (Druin, 2002; Massion & Nam, 1998). Knowing that “no amount of adult evaluation will find all the issues that children will stumble into,” it becomes important to
involve representative children in the usability tests (Hanna et al., 1997, p. 14). In fact, this is considered a main principle in user-centered design (Barendregt et al., 2007; Markopoulos et al., 2008).

**Challenges of usability testing with children.** The literature has specified several factors that could affect the process of usability testing and that require the designer’s attention when designing usability tests. These factors can make the usability testing process more challenging and demanding. One of these factors is that a small difference in children’s age makes a big difference in attributes. In other words, “there may be just as big a difference between second and third graders as there is between first and fifth graders” (Byerly, 2007, p. 32). This fact requires designers to conduct the usability tests with a large number of users in order to cover the usability of the product as perceived by different ages.

Another factor is children’s short attention span and their tendency to be distracted and to feel bored (Hanna et al., 1997; Rounding et al., 2013). When children are distracted or bored it is important to know how to gently bring them back to the tasks of the test and when to give them a break (Hanna et al., 1997). However, the test moderator should never force the child into continuing the session. This unethical act might distress the child. In such cases the test session must be stopped promptly (Markopoulos et al., 2008).

Another factor that is also related to ethical issues is that it is always advisable to conduct the test with the presence of the parent or at least another adult while recording the whole session (Markopoulos et al., 2008). Even though this might reduce the effectiveness of the test and/or increase its costs, it is essential because it will protect the
child from child abuse, and it will protect the evaluator from being accused of child abuse. Other ethical concerns that add more challenges to usability testing with children are maintaining the safety of children, collecting the parent’s consent, and making sure that everything in the process is clear and agreed upon (Khanum & Trivedi, 2012).

Children’s language abilities can also influence the testing results (Markopoulos & Bekker, 2003; Markopoulos et al., 2008). Therefore, the designer should pay close attention to the words and terminology used with the children (Byerly, 2007; Markopoulos et al., 2008). Children’s language abilities do not only affect their understanding of the questions, but also the quality of their answers. For example, a child might not be able to give the answer he or she is thinking of because of the difficulty of finding the right words. Hence, designers should take into account the type of answers required from the child based on the child’s abilities and age. For example, instead of asking the child to write or talk about his feelings, the picture card method could be used where the child could pick from a set of smiley faces cards that express the different feelings the child might have (Markopoulos et al., 2008).

Another factor that could affect usability testing is that some children have trouble in thinking aloud (Barendregt et al., 2007; Khanum & Trivedi, 2012). They might feel uncomfortable to talk about what they are thinking while using the product (Khanum & Trivedi, 2012). This might be due to the high cognitive load posed by this method. Sometimes they simply forget to think aloud (Barendregt et al., 2007). It is not advisable, though, to keep prompting them to think aloud when they forget because this could lead them to mention non-existing problems just to please the evaluator (Barendregt et al., 2007).
Barendregt et al. (2007) have investigated the children’s personality characteristics that can predict the effectiveness of usability testing, especially the thinking aloud method. They found that selecting children with high curiosity scores led to the detection of the highest number of problems, while selecting children with high extraversion scores and low friendliness scores led to the detection of the highest number of problems through self-initiated verbalization. Therefore, it might be possible to screen the participants based on these characteristics, but it is generally advisable to avoid screening in order to reduce the workload for the children and their parents (Markopoulos et al., 2008). In all cases, if the participants of a usability test are selected from a classroom, it is unethical to pick some of the students based on their characteristics. In such cases, the test moderator has to conduct the test with all students and discard the cases that are irrelevant or non-representative (Markopoulos et al., 2008).

Another challenge is the difficulty of adjusting to unfamiliar settings (Markopoulos & Bekker, 2003). This factor might increase the child’s distraction during the test. Therefore, it is important to take the children in a tour around the building and to show them the testing room and the testing equipment in order to help them decrease their fears (Hanna et al., 1997; Markopoulos et al., 2008). Another strategy is to keep the parent present during the test, but it is important in this case to prevent the parent from affecting the child’s answers or performance (Hanna et al., 1997). Another solution to the problem of unfamiliar environments is to conduct the test in a more familiar place, such as the child’s school or at the child’s home. This can be through different methods such as naturalistic observation, where the test moderator passively observes the use of a product in a classroom, or through field evaluation, where the child keeps a diary at
home, for example (Markopoulos et al., 2008).

Although conducting the usability test at home or school can be a solution to the difficulty of adjusting to new environments, it can also be problematic for the parents and the teachers. As mentioned by Druin (2002), children’s participation in usability testing can also be challenging for parents and teachers as well. Teachers may need to spend class time on testing learning products rather than teaching. If the test is being conducted outside of school, parents may need to take their children to a lab, or they may need to allow the researchers to come to their home; “this necessitates time and energy, which many of today’s parents have little” (Druin, 2002, p.14). Markopoulos et al. (2008) also mentioned that it is challenging to find parents who agree to host usability tests in their home. Crowdsourced or online usability testing might be beneficial in such cases since it is typically conducted at home but without the presence of the researcher or the product developer. This solution may not only help the child and parents feel more comfortable, but it may also help the designer to obtain more realistic results since the environment of the test is similar to the real-life use of the product. On the other hand, it does not provide the same level of control over the testing situation (Markopoulos et al., 2008).

Children’s logical thinking is also one of the factors that add to the challenges of usability testing with children. Because the logical thinking of children is not mature, it is sometimes hard for them to compare products and provide reasoning or even understand abstract tasks (Markopoulos & Bekker, 2003). Therefore, it is important to adapt the usability testing methods to make them suitable for the children’s abilities. The immaturity of logical thinking and reasoning abilities requires the designer to triangulate the data collected by using more than one testing method. Using different methods and
techniques will allow the designer to record the child’s opinions, thoughts, and emotions in different ways that do not rely heavily on a single ability, such as verbalization (Khanum & Trivedi, 2012).

It has been shown in this section that the challenges of usability testing with children emerge from different aspects, including the child’s characteristics, the test’s environment, the test’s methods, the wording of the test’s tasks and instructions, the ethical practices during the test, and the inconveniences imposed on parents and teachers. Careful preparation is the key to diminish the effects of most of these challenges. “When designing test tasks, children's language, cognitive abilities, and developmental levels must be taken into account” (Markopoulos et al., 2008, Chapter 6, Section 12, para. 2). After designing the test tasks based on these considerations, pilot testing has to be conducted in order to ensure the appropriateness of the test set-up, the clarity of the tasks, and the minimization of negative effects on children users (Markopoulos et al., 2008).


**Recording and logging methods.** Recording and logging methods can be used with any of the other usability testing methods. They include the automated recording of the child’s interactions with the user interface, which are usually recorded in video files of the screencast in addition to spreadsheets or text files. They also include video and audio recordings, which can be used to record the child while being interviewed or while
using the product. Eye tracking is also considered a recording method. It provides the designer with valuable information on the user interface by providing a heat map which specifies which areas of the interface were ignored and which were looked at based on tracking and analyzing the child’s eye movement and fixation (Byerly, 2007). The downfalls of this method are its high costs (Markopoulos et al., 2008) and the complexity of the heat maps’ coding and interpretation (Byerly, 2007). There are several ethical issues related to the recording and logging methods. These issues will be discussed later in this chapter.

**Observation methods.** The second type of usability testing methods with children is the observation methods. Several decisions have to be made in this type. The designer needs to decide whether the observation is going to be in the field or in the lab; direct or indirect (Khanum & Trivedi, 2012); participant or passive; and if it is going to be structured or unstructured (Markopoulos et al., 2008). These decisions might affect the reliability, realism, and the ecological validity of the results. For example, unstructured observations can reduce the reliability of the results, while passive and field observations can increase the realism of the testing situation (Markopoulos et al., 2008). The decision, therefore, should rely on evaluating the designer’s options and resources and selecting the optimum method.

**Verbalization methods.** The single type of methods that has obvious differences between testing with children and testing with adults is the verbalization methods, which result in “a record of things testers say during usability testing” (Markopoulos et al., 2008, Chapter 10, Section 1, para. 3). One of the verbalization methods is thinking aloud, which is the most widely used method by usability testing practitioners (Khanum
In general, there are several advantages of the think aloud method (Nielsen, 2012b); it is cheap, robust, flexible, convincing, easy to learn, and most importantly, it shows what the user actually thinks about the product. The disadvantages of the think aloud method is that it does not provide statistical results; it does not reflect the realistic use of the product; the user’s thoughts are sometimes filtered before talking; the user’s behaviors are sometimes biased by the test administrator interference; and like all the other methods, think aloud should be used with other usability testing methods in order to achieve all purposes of usability testing (Nielsen, 2012b).

Markopoulos et al. (2008) shed the light on the differences between using think aloud with children and adults. In adult’s think aloud, the test administrator is required to maintain the least interference with the person thinking aloud in order to ensure the validity of the verbalized data. As Nielsen (2012b) describes it, the designer needs to do only three things in order to conduct a basic think aloud usability study: “1- Recruit representative users. 2- Give them representative tasks to perform. 3- Shut up and let the users do the talking” (para. 4). On the other hand, Markopoulos et al. (2008) demonstrate that there are several factors that require the test administrator to do the opposite when using think aloud method with children. These factors can be summarized as follows: children’s language skills prevent them from expressing their thoughts in a clear manner; children are more likely to not think aloud when facing a problem due to the high cognitive workload at these moments; and the unfamiliarity of the product and the test administrator makes children feel shy and refrain from talking, especially if they think that there is a right or wrong answer or that they are being tested and not the product.
Therefore, “related texts suggest a more dialogical demeanor by the administrator, which allows more social interaction and more guidance to offer” when conducting think aloud usability studies with children (Markopoulos et al., 2008, Chapter 10, Section 1, para. 3). Following this approach allows the think aloud method to be used with children as young as seven (Markopoulos et al., 2008), rather than limiting its use to older children (Hanna et al., 1997). It can be said that the context and the general atmosphere affects the children’s ability to think aloud. In a study by Khanum and Trivedi (2013b), it was found that testing children in the field rather than in the lab resulted in better verbalization quality.

The Wizard of Oz method. Another type of usability testing methods with children is the Wizard of Oz, which, according to Markopoulos et al. (2008), was introduced by Gould, Conti, and Hovanyecz in 1983. This method is used when it is expensive to produce the product under investigation. Therefore, the designers produce an incomplete prototype or simulation that looks like a complete product. The difference is that a human wizard would be responsible for the interactive parts of the product in a way that does not make the child feel that a human is controlling the interactions. While the child is using the prototype on one computer, the human wizard will be watching the child’s interactions with the prototype through a screen in a different room in order to make the prototype respond accordingly. For example, if the tested product is supposed to respond verbally to the child’s questions, the human wizard will hear the child’s question from another room and respond to the child’s question while the child thinks that the answer is coming from the product itself. The advantage of using this method is that it reduces the costs of producing expensive products.
**Survey methods.** Survey methods need great adaptation when used with children. They include questionnaires, rating scales, and structured interviews. There are several points that need to be taken into account when designing usability surveys for children (Markopoulos et al., 2008). The designer needs to understand that some children might not be able to read the questions properly due to their reading age. On the other hand, when a child is able to read a question, it does not mean that he or she is able to understand it. The quality of a child’s responses is not only affected by the possessed language abilities, but also by the motor abilities and temperamental effects. Therefore, designing proper survey questions for children is not an easy task. One of the ways to ensure simplicity and clarity of children’s survey questions, besides conducting pilot studies, is to ask a teacher to review the questions (Markopoulos et al., 2008).

Unfortunately, there is a lack in the literature on the effectiveness, the validity, and reliability of children’s survey methods and responses (Markopoulos et al., 2008).

Visual Analogue Scales (VAS) is a commonly used question format in children’s questionnaires that were designed to elicit the child’s opinions or feelings (Markopoulos et al., 2008). A VAS scale consists of faces that start with a happy face and gradually change to end with a crying face. The problem with this kind of scales when used to elicit children’s opinions about a computer product is that the younger children tend to always choose the highest score (Read & MacFarlane, 2006). Hanna, Risden, Czerwinski, and Alexander (1999) have developed the Funometer scale for measuring the usability and engagement in children’s computer products. It consists of a vertical scale that has a smiling face and a frowning face at its two ends. The child can draw a line...
from the frowning face to the smiling face to indicate how much fun was the activity or the product (Markopoulos et al., 2008).

Read and MacFarlane (2006) extended the Funometer scale by designing the Fun Toolkit, which consists of four tools that are useful in Child Computer Interaction survey studies. The four tools are the Funometer, the Smileyometer, the Again – Again Table, and the Fun Sorter. The Smileyometer looks like a Likert scale but with smiley faces, and it is used before and after a task to measure expectations and judgments. It is not recommended to use the Smileyometer with children younger than seven years old because it is expected that they would tend to choose the highest score on the scale (Read & MacFarlane, 2006). The Fun Sorter can be used to rank the tasks or products based on specific constructs, while the Again – Again table can be used to measure the engagement of the children in a specific activity or product by asking them whether they would do the activity or use the product again.

After validating the Fun Toolkit, the researchers found that the Again – Again table has more validity than the Smileyometer when used with very young children. The Again – Again table was also found to be better than the fun sorter in which the children tend to change their answers in order to be fair with all the activities or the products and not to make one of them take the worst score in every construct (Markopoulos et al., 2008). They also found that the Funometer can be used to measure the ease of use instead of using it to measure the fun construct, which is also measured by the Smileyometer and the Again – Again table (Read & MacFarlane, 2006).

**Diary method.** Since the diary method is a long-term usability evaluation that is conducted in the field without the presence of a test administrator, it is therefore
considered as the only method that represents the actual use of the product (Markopoulos et al., 2008). In other words, the realism and the ecological validity of this kind of usability test are higher than in the other methods. A diary can be described as “a form of survey in which a diarist, the participant in such studies, independently answers questions set by the evaluator over a sustained period of time” (Markopoulos et al., 2008, Chapter 14, Section 2, para. 1).

The parent evaluator method, as described by Markopoulos et al. (2008), is a diary method that is carried at home, and it was originally designed for younger children. Testing at home with the help of the parent is useful in that it provides a familiar environment for the child; children are more comfortable answering parents’ questions rather than talking to a stranger test administrator; and the testing situation is similar to real life situations in which the product is introduced by the parent. However, there are some disadvantages like the need to train the parents on their role and to make them behave based on the idea that the product is what is being tested and not their children. The parent’s role is maximized in this method and it includes the roles usually taken by the test administrator, the observer, the technical troubleshooter, and the diarist. Therefore, it is recommended to find ways that reduce the workload of the parent in this method (Markopoulos et al., 2008).

**Inspection methods.** These methods rely on usability experts in evaluating the usability of the product. They do not involve testing the product with real users, and therefore they cut the costs of usability testing labs and human resources (Markopoulos et al., 2008). On the other hand, inspection methods might lead to misleading findings since children users are not actually involved in the process. As discussed earlier, adult
evaluators cannot predict all the usability problems that will be faced by children (Hanna et al., 1977; Markopoulos et al., 2008). Inspection methods are criticized for their subjectivity and lack of reliability (Khanum & Trivedi, 2012). However, when inspection methods are used with other usability methods, it increases the detected number of usability problems because it allows the product to be checked from different viewpoints and perspectives (Markopoulos et al., 2008; Wilson, 2007).

In fact, the use of variant usability testing methods is not only important in increasing the detected number of usability problems, but it is also advisable for several reasons. It allows for triangulation of data, which helps in confirming the evaluators’ conclusions (Markopoulos et al., 2008). It can be also used to serve different purposes of usability testing, since a single method does not serve all purposes (Nielsen, 2012b). It is even advisable to apply lab and field methods with children testers in order to obtain real results (Massion & Nam, 1998). Finally, relying on a single usability testing method is specifically undesirable with children users because they have limited and varied levels of capabilities (Khanum & Trivedi, 2012). This factor limits the benefits of using a single method, which usually depends on one of the child’s capabilities of expressing their thoughts and opinions.

**Lab versus field in usability testing with children.** The decision of using a specific usability testing method is related to the decision of whether to conduct the test in the field or in the lab. Therefore, it is not only important to know what different methods can offer, but it is also important to understand the advantages and the disadvantages of testing in the lab versus testing in the field. This section summarizes the most important comparison points between the two settings when testing with children.
Testing in the lab provides more control over the situation. For example, unlike testing at home, distractions by the siblings of the child can be controlled during the test at a lab (Hanna et al., 1997). However, studies have found that interruptions during the test in the field setting did not influence the children’s performance, and that there was no significant difference between the problems detected in both settings (Khanum & Trivedi, 2013a). Testing in the lab also has the advantage of “team involvement in observing children” (Hanna et al., 1997, p.10). On the other hand, testing in the field has the advantage of using devices that the children are familiar with (Hanna et al., 1997). Therefore, testing in the field does not require the child to face any difficulties switching to unfamiliar input devices or to adapt to a new environment like in the lab setting (Markopoulos et al., 2008). It provides higher ecological validity since it simulates the actual use of the product and facilitates the detection of real usability problems (Markopoulos et al., 2008). Being tested in a familiar environment, children were found to be more relaxed in the field than in the lab (Khanum & Trivedi, 2013a).

Testing in the field costs less in terms of money and time (Khanum & Trivedi, 2013a). It was also found that children revealed more severe problems when performed the test in the field than when performed it in the lab (Khanum & Trivedi, 2013a). In a study that compared the quality of the verbalized data between lab and field tests, it was found that children testers provided better verbalization in the field (Khanum & Trivedi, 2013b). The presence of the parent or teacher in the field adds to the value of the testing results since they know the child and therefore can put the child’s responses or performance into context (Markopoulos et al., 2008). Conducting a usability test at children’s school is easier than bringing them to the lab. However, the cooperation of the
teachers is substantial for the success of the process (Markopoulos et al., 2008). It is problematic for the teachers to spend their class time on testing learning products rather than investing it in teaching and learning (Druin, 2002). On the other hand, a disadvantage of usability testing at the child’s home is the difficulty of finding “participants who are willing to host a usability test in their own home” (Markopoulos et al., 2008, Chapter 5, Section 12, para. 2).

Based on these points, testing in the field seems to be more convenient with children users. Whether the usability test is going to be conducted in the field or in the lab is a decision needs to be made by the designer. In both situations, the designer needs to understand the influence of the adults’ presence during the test.

**Adults’ influence.** Whether the adult is a test administrator, a parent, or a teacher, his or her presence during a usability test might affect the child’s performance and responses. One reason is that some children feel shy or uncomfortable around strangers, especially when they think that they are being tested and not the product. Being uncomfortable affects their ability to verbalize their thoughts, or it leads them to mention problems that do not exist just to please the observer (Markopoulos et al., 2008). Van Breemen, Yan, & Meerbeek (2005) developed Philips iCat that can replace the presence of the human test administrator with the child in one room. Philips iCat is “a desktop user-interface robot with mechanically rendered facial expressions” (van Breemen, Yan, & Meerbeek, 2005, p. 143). iCat was used in several usability testing studies in which the test administrator watches the child user through a one-way mirror and a camera while interacting with him or her through iCat’s facial expressions, head
movement, and built-in microphone (e.g. Fransen & Markopoulos, 2012; Markopoulos et al., 2008; van Breemen et al., 2005; Verschoor, 2007).

Studies that used iCat found that children were more comfortable talking to the robot than talking to the test administrator, but there was no difference between the amounts of verbalized data in both situations (Fransen & Markopoulos, 2012). The children were having fun and became more engaged once the robot woke up and started interacting with them (Fransen & Markopoulos, 2012; Verschoor, 2007). It was also found that controlling iCat during the test increased the workload of the test administrator, which did not allow for the detection of major usability problems during the session. However, the built-in camera helped in watching the video of the test later after the session was done (Verschoor, 2007). After several tryouts and iterations, Fransen and Markopoulos (2012) provided a list of guidelines for the application of social robots in the usability testing with children.

**Guidelines of Usability Testing with Children**

Although several researchers have attempted to provide some guidelines for the application of usability testing with children, most of these guidelines are narrow and directed to specific methods. The most comprehensive list of guidelines was provided by Microsoft’s usability engineers, Hanna, Risden, and Alexander (1997), and it is still applicable and being used in usability studies today (e.g. Barendregt & Bekker, 2003; Khanum, & Trivedi, 2012; Khanum, & Trivedi, 2013a; Khanum & Trivedi, 2013b; Markopoulos & Bekker, 2002; Markopoulos & Bekker, 2003; Meloncon et al., 2010). In fact, it is hard to find a study of usability testing with children that does not at least
mention Hanna et al.’s contribution to the field. The limitation of this set of guidelines is that it was mainly developed for lab-based usability-testing sessions.

Hanna et al. (1997) have specified the appropriate age range for these guidelines as between 2½ and 14 years old, as younger children are “not proficient enough with standard input devices,” and older children “will likely behave as adults in a testing situation” (p. 10). This statement might not be accurate after about two decades of the development of these guidelines, especially with the advancement in input devices and the level of experience children have with the technology nowadays. Hanna et al. divided the guidelines into four sections that represent the stages of the usability testing process: Set-up and planning, introductions, during the test, and finishing up. The following table lists the guidelines of each stage. It also includes the researcher’s initial adjustments to the wording of each point to make it suitable for online usability testing since some points in the original version of the guidelines are not applicable for the online approach, which is the main focus of the current study. The changes did not affect the main ideas of the guidelines, which were investigated in the current study in order to create a new list of guidelines that are suitable for online usability testing.
Table 1

*Hanna’s Guidelines of Usability Testing with Children*

<table>
<thead>
<tr>
<th>Hanna’s Original Guidelines (Hanna et al., 1997)</th>
<th>The Researcher’s Initial Adjustments to Hanna’s Guidelines for Online Usability Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Set-up and planning</strong></td>
<td></td>
</tr>
<tr>
<td>The lab should be decorated in a manner that would attract the children but not distract them.</td>
<td>Parents should be instructed to eliminate any distractors in the room where the test is taking place, and the child’s interface of the usability testing website should be attractive but not distracting.</td>
</tr>
<tr>
<td>The input devices used in the lab should be the same as the ones they use at home or school.</td>
<td>N/A</td>
</tr>
<tr>
<td>Testing equipment should be selected and placed wisely and unobtrusively in the lab to capture the optimum sound and picture in a least-distracting manner.</td>
<td>Recording tools, including on-screen tools, should be placed wisely and unobtrusively to capture the optimum sound and picture in a least-distracting manner.</td>
</tr>
<tr>
<td>Fewer children participants should be scheduled on a usability-testing day than it is usually done with adult</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Table 1 continued

<table>
<thead>
<tr>
<th>Topic</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>If a testing session is supposed to take 30 minutes, a child should be scheduled for an hour; children tend to spend some time playing and exploring.</td>
<td>If a testing session is supposed to take 30 minutes, a child in an online usability testing should be scheduled for an hour; children and parents will need the extra time in reading the instructions and practicing before beginning the actual test.</td>
</tr>
<tr>
<td>Tasks should be reordered with each participant so that the children’s experience and fatigue do not affect their performance on the same tasks.</td>
<td>The instructional designer should have the option of reordering the tasks automatically with each participant so that the children’s experience and fatigue do not affect their performance on the same tasks.</td>
</tr>
<tr>
<td>Children should have at least six months of experience with the input devices.</td>
<td>Children should have at least six months of experience with the input devices.</td>
</tr>
<tr>
<td>In order to avoid bias, it is better not to use one’s own children or a friend’s children as participants in the tests.</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**Introductions**

A relationship with the child should N/A
Table 1 continued

be established by starting the test
with a small, friendly conversation.

The importance of keeping what the
child and the parent will see or know
about the design of the product as a
secret should be discussed with
them.

The importance of keeping what the child
and the parent will see or know about the
design of the product as a secret should be
included in the agreement at the beginning
of the usability test.

Introducing the children to the
testing situation should be done in
the same way for all participants by
following a prepared script.

An introduction video should be used in
order to ensure that introducing the children
to the testing situation is done in the same
way for all participants.

In order to motivate older children
for better cooperation, they should
be told how important their role is in
making a good product for other
children.

In order to motivate older children for better
cooperation, they should be told how
important their role is in making a good
product for other children. This information
should be included in the introduction video.

If a low-fidelity prototype is used in
the test, the importance of using such
prototypes in the development
process of the product and the

N/A
Table 1 continued

<table>
<thead>
<tr>
<th>Importance of the child’s role in this stage should be explained.</th>
<th>A tour in the lab should be given to the child and the parent with an explanation of what will be happening and why the session is going to be recorded.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A website tour should be provided as a video with an explanation of what will be happening and why the session is going to be recorded.</td>
<td>Younger and shy children should have their parents stay with them during the sessions, while the parents of older children should be allowed to watch the session from another room. Parents should be instructed to support their children when needed and to minimize their interaction with them.</td>
</tr>
<tr>
<td>While an observation of older children can be done from another room while sending softly spoken instructions through the speakers, it is recommended that the test administrator stay in the testing room.</td>
<td>Parents should stay at the testing room with the child during the session since homes are not usually prepared with one-way mirrors or cameras and monitoring devices that allow them to stay in another room.</td>
</tr>
</tbody>
</table>
Table 1 continued

with younger children to encourage them and to minimize parents’ interference.

Siblings should not be allowed to stay with the child participant in the testing room to reduce distraction and interference. The instructions should include that siblings are not allowed to stay in the testing room during the test to reduce distraction and interference.

During the test

With younger children, a warm-up activity should help them become familiar with the input devices used in the session. With younger children, a warm-up activity should be used to help them become familiar with the input devices. For example, the child can spend a few minutes playing a game that helps in using the mouse or the touch pad.

While both younger and older children can perform free exploration, it is possible for older children to perform short tasks as well. However, it is important to understand the limits of children. They cannot perform long and complex tasks, they might forget the

Long tasks should be broken down into shorter tasks, parents should be instructed to restate the task when the child forgets it, and a list of hints should be prepared and provided to the parent to support the child when needed.
Table 1 continued

- task, and they might need some hints. Therefore, the tasks should be broken down, they should be restated when the child forgets them, and a list of hints should be prepared to support the child when needed.

- When children ask a question on how to perform a specific task, their question should be redirected by asking them how they think they can perform it or by asking them to guess.

- Children should be instructed to perform the tasks, but not asked if they want to do them. Otherwise they can say no.

- Children should be gently reminded to complete the tasks when they are distracted, and they should be offered a short break when the session is longer than 45 minutes.

- Parents should be instructed on the right way to answer questions during the test.

- When children ask a question on how to perform a specific task, the parent should redirect their question by asking them how they think they can perform it or by asking them to guess.

- Parents should be instructed not to ask the children if they want to perform the task, but to directly instruct them to perform it.

- Parents should be instructed to gently remind the children to complete the tasks when they are distracted, and they should offer them a short break when the session is longer than 45 minutes.
Table 1 continued

| A good strategy to engage the child in a task is to pretend that his or her help is needed to complete the task. | A good strategy that parents need to be aware of when a child is not engaged in a task is to pretend that his or her help is needed to complete the task. |
| Written items should be read for children who are not proficient enough in reading, if it is not part of the task itself. | Written items should be read for children who are not proficient enough in reading, if it is not part of the task itself. This can be added as an option in the website or the parents should be instructed to read the items for the children. |
| When children struggle in figuring out the product, it is important to provide positive feedback that does not indicate whether they completed or failed to complete the required tasks. | When children struggle in figuring out the product, it is important for the parents to provide positive feedback that does not indicate whether they completed or failed to complete the required tasks. |

**Finishing up**

| Body language, face expressions, and behaviors should be all used besides the child’s responses because they are more reliable and can indicate whether the child is engaged. | Body language, face expressions, and behaviors should be all used besides the child’s responses because they are more reliable and can indicate whether the child is engaged or disengaged more than his or her |
or disengaged more than his or her responses alone. These signs can be recorded using a built-in camera or can be observed and reported by the parent.

Older children can provide information on how to improve a product, and their responses to rating scales such as the Funometer can actually be reliable.

Gratitude should be expressed to the children. A payment or an electronic gift should be offered for children and parents as a gratuity at the end of the testing session.

“For a gratuity, children and parents often appreciate a choice of a gift certificate to a local toy store or movie theater in addition to standard software or payment options”

(Hanna et al., 1997, p. 14).

Changes have occurred in technology and children’s use of technology since Hanna’s guidelines were published more than 20 years ago. According to Childwise 2015 report, UK children’s average time spent on screen increased from three hours in 1995 to six and a half hours in 2015 with a decrease of TV watching time (Childwise, 2015). In a sample of 2000 children between 5 and 16 years old, an evident increase has been noticed in computer and tablet ownership from 2013 to 2017, especially among
younger children. Fifty-five percent of children 5 to 11 years old had their own computer in 2013 compared to 80% in 2017 (Childwise, 2018). Although researchers still use Hanna’s guidelines in usability testing studies, it is important to recognize that changes in children’s use of technology probably require changes in the guidelines of usability testing with children.

It is worth mentioning that other usability evaluation experts, beside Hanna, Risden, and Alexander, have developed guidelines for usability testing with children. However, Hanna et al.’s guidelines were the most comprehensive, while most of the other guidelines were more specific to a single method and/or age range. Read and MacFarlane (2006), for example, have suggested a general set of guidelines for children’s survey methods. Barendregt and Bekker (2003) have attempted to extend Hanna et al.’s guidelines by adding some suggestions for computer games usability testing with children of five to seven years old. Markopoulos et al. (2008) have also provided several sets of guidelines for different methods of children user tests. Generally speaking, any set of guidelines has to take into consideration the ethical issues surrounding usability testing with children. These considerations will be discussed in the following section.

**Ethical Considerations**

Usability testing with children involves several ethical concerns (Khanum & Trivedi, 2012; Markopoulos et al., 2008). It is important to understand these concerns in order to mitigate the risk of falling into ethical holes. Markopoulos et al. (2008) clarifies that usability testing may lead to physical or psychological harm if not carefully planned and implemented. Pilot testing can be helpful in reducing these kinds of risks. Parents’ permission has to be collected in the form of informed consent. It is also important to
make sure that the child accepts to participate in the study by explaining it to him or her (Markopoulos et al., 2008), emphasizing that the product is what is being tested (Nielsen, 1993b).

The child needs to understand that he or she can opt out of the study at any time without consequences (Markopoulos et al., 2008). Children might have difficulty expressing their desire to opt out, but it can appear in their body language or behaviors. The test administrator must notice these signs and ask the child if he or she wants to withdraw (Markopoulos et al., 2008). In such cases, children should still receive the incentive promised by the test administrator, even if they did not perform any tasks; all participants who “take time out of their schedules, show up, and contribute to your study deserve the compensation whether or not they finish” (Sova & Nielsen, 2008, p. 46).

Confidentiality of the test results and the privacy of the users’ identities should be maintained (Nielsen, 1993b). If the test involves using web content, two issues need to be taken into consideration; it is important to protect the child from being exposed to inappropriate content and to protect him or her from sharing personal information (Markopoulos et al., 2008). The child has the right to know if he or she will be photographed or video/audio recorded, and permissions should be collected from the child and the parent (Markopoulos et al., 2008). Sharing video recordings outside the usability testing team is considered unethical (Sova & Nielsen, 2008). Permission has to be collected if the video recording is going to be shared with any person outside the usability testing team (Nielsen, 1993b). In fact, this rule does not only apply on video recordings, but on audio recordings as well, since a child can be identified from his or her voice (Markopoulos et al., 2008).
Special attention should be given to the difficulty and the length of the tasks in order to avoid distressing the child (Markopoulos et al., 2008). The experimenter is responsible “to make the users feel as comfortable as possible during and after the test. Specifically, the experimenter must never laugh at the users or in any way indicate that they are slow at discovering how to operate the system” (Nielsen, 1993b, p.182). In case a child became distressed for any reason, the test administrator should stop the session immediately (Markopoulos et al., 2008).

It is a good idea to prepare a script to be used with the children in the introductions stage of the usability testing session. This script will ensure that the test administrator will not forget to explain the important points, such as the right to opt out and the purpose of the video recording (Markopoulos et al., 2008). However, it is important to read the script in a natural way in order to help the children feel comfortable (Barendregt & Bekker, 2003).

Finally, several scholars have recommended the use of the general code of ethics of some associations, such as the American Psychological Association (APA) and the Association of Computing Machinery (ACM) (Markopoulos et al., 2008; Sova & Nielsen, 2008). These ethical principles are not specific to children, but they can be a good “starting point” (Markopoulos et al., 2008).

**Summary**

In this chapter, a literature review has covered the main topics of the present study, including the cognitive load theory, usability testing, and crowdsourcing. Traditional usability testing poses several financial, human resources, and other challenges. It was shown that crowdsourcing has promising potentials that can be used to
overcome these challenges. It was also shown that crowdsourcing is already being used with adult testers, but has not been applied or tried yet on children users. Usability testing with children is demanding and often discarded. Some designers tend to test children’s products with adult testers or experts. This could lead to products with severe problems that could hinder their purposes. A new usability testing method like crowdsourcing might be helpful for testing children’s products with children. Several considerations need to be taken into account when designing such tools. New guidelines for such methods are needed and ethical practices need to be maintained for the safety of the children and the designers.
Chapter 3: Methodology

The purpose of this study is to develop practical guidelines for the design of usability testing for children’s products in an online setting. Although many websites have been developed for usability testing, the researcher could not find any websites that allow designers to test children’s products with children users. Therefore, the focus of this study is to develop a set of guidelines for this kind of testing in order to facilitate future research and development of online usability testing with children. The research questions are:

1. What can be learned about the remote usability testing guidelines in terms of:
   a. the features and characteristics of the usability testing tool that are related to the children nine to ten years old?
   b. improving the role of parents in such a way that can reduce the workload of parents while maintaining a good quality of usability testing results?

2. What concerns do parents have regarding their children’s participation in crowdsourced usability testing?

These questions were answered using a design and development research methodology.

Design and Development Research Methodology

A design and development research is defined by Richey and Klein (2014a) as “the systematic study of design, development and evaluation processes with the aim of establishing an empirical basis for the creation of instructional and non-instructional products and tools and new enhanced models that govern their development” (p. 142). It is different than design-based research, also called educational design research, which is defined as “a genre of research in which the iterative development of solutions (e.g.,
educational products, processes, programs or policies) to practical and complex educational problems, provides the setting for scientific inquiry, and yields new knowledge that can inform the work of others” (McKenney & Reeves, 2014, p. 133).

Both types of research overlap in the research projects that actively solve real problems in educational practice. However, the research projects that fall exclusively under the design-based research are the ones that are concerned with instructional products or specific situations. Similarly, there are research projects that are exclusive to the design and development research; those are the ones “concerned with developing tools or models to support education in the long run but that do not function as educational interventions (e.g. retrospective analysis of how instructional designers carry out their tasks)” (McKenney & Reeves, 2014, p. 133). Richey and Klein (2014a) have been advocating this type of research in the instructional design field for almost two decades. They confirm that there is a critical need for this type of research especially since only “few models, design strategies, and tools employed in practice have been empirically tested and validated” (Richey & Klein, 2014a, p. 142).

Richey and Klein (2014a) divide this genre of research into two major types: Type 1, which is research on products and tools, and Type 2, which is research on design and development models. Type 1 research is conducted during the product’s design and development. Its outcomes are the “lessons learned from developing specific products and analyzing the conditions which facilitate their use” (Klein, 2014, p. 3). In this type, either the whole process or part of it is documented (Richey & Klein, 2014a). Type 2, on the other hand, is concerned with developing and validating design and development models and their use (Richey & Klein, 2014a); its outcome is a new design model (Klein,
While the research on models is generalizable, the research on products and tools is contextually specific (Klein, 2014).

The current study falls under Type 1: Research on products and tools. Richey and Klein (2014a) specified four phases as a framework for the design and development of products and tools: analysis, design, development, and evaluation. The analysis phase has several elements including the statement of problem, literature review, and research questions (Herrington, McKenney, Reeves, & Oliver, 2007). The design’s assumptions are formed based on the outcomes of the analysis phase and are applied in the prototype development phase. The prototype development in this study is primarily based on the cognitive load theory and on Hanna et al.’s (1997) guidelines of usability testing with children. The prototype was first tested in a pilot study, which helped in reaching appropriate test set-up, clear tasks, and minimized negative effects on the users (Markopoulos et al., 2008), and then it underwent a series of iterations as will be shown shortly. The implementation of the four phases in the current study will be discussed in the following section.

**Design and Development Research Study Design**

**Analysis.** “The value, authenticity, and the perceived relevance of a design and development research project depend to a great extent upon the problem selected” (Richey & Klein, 2014b, p. 16). The problem in the present study was identified through the current research and literature. It was found that the traditional usability testing methods pose a number of challenges, including: The difficulty to recruit testers who have the same characteristics as the typical users of a product, the high costs of the tests, and the length of the process (El-Halees, 2014; Harty, 2011; Moore, 2008; Rubin &
Accordingly, new approaches are needed to improve the process of usability testing, especially when testing children’s products. Usability testing with children can be more challenging in comparison to usability testing with adults due to differences in experience and in cognitive, physical, and social development (Khanum & Trivedi, 2012; Markopoulos et al., 2008).

The literature indicates that crowdsourcing is currently used as a new approach to overcome some of the usability-testing challenges. However, after analyzing a number of crowdsourced usability testing websites, it was found that none of these websites focused on the appropriate testing of children’s interactive products. In fact, adult users are recruited to test all kinds of products, whether they were designed for adults or children. Given the challenges of traditional usability testing methods and the importance of recruiting representative users, there is a need to investigate the feasibility of crowdsourced usability testing with children and to explore its benefits. Therefore, it is important to start with finding a set of practical guidelines for the design of this approach considering the unique characteristics of children, the preferences of the parents, and the capacity of online environments.

The objective of the present study is to create a set of guidelines for the design of online usability testing websites for children’s interactive products. Creating a set of guidelines is the first attempt to investigate the design of remote usability testing of children’s interactive products. In other words, extended research on the use of crowdsourced usability testing with children will be easier with a set of guidelines for the design of such platforms. Such studies will eventually provide the instructional designers
of children’s products with a solution to the challenges of usability testing, which were discussed in the earlier chapters, through the use of crowdsourcing.

The guidelines focus on obtaining sufficient usability testing results that allow the instructional designers to properly improve their products while reducing the workload of parents who monitor the remote usability test through a crowdsourcing website. In general, this study aims to identify how we can best take advantage of the potential of crowdsourcing in the usability testing of children’s interactive products by taking the right design decisions.

**Design.** A prototype of the crowdsourcing usability testing website, Kid Crowd, was designed in the current study based on Hanna et al.’s guidelines (1997) and the principles of cognitive load theory. Although several researchers have suggested different guidelines for usability testing with children, Hanna et al.’s (1997) guidelines were chosen because they are the most comprehensive list of guidelines in the literature, and they are still applicable and widely used in usability studies that involve children’s products (e.g. Barendregt & Bekker, 2003; Khanum, & Trivedi, 2012; Khanum, & Trivedi, 2013a; Khanum & Trivedi, 2013b; Markopoulos & Bekker, 2002; Markopoulos & Bekker, 2003; Meloncon et al., 2010). As shown in Chapter 2, Hanna et al.’s guidelines are divided into four stages that represent the stages of laboratory-based usability testing: Set-up and planning, introductions, during the test, and finishing up. Each stage consists of a set of guidelines that were identified for children between 2½ and 14 years old (Hanna et al., 1997).

To comprehend what functions a crowdsourced usability testing website should include, the researcher investigated three crowdsourcing books. These books are *The
The wisdom of crowds, by James Surowiecki (2005); Crowdsourcing: Why the power of the crowd is driving the future of business, by Jeff Howe (2008); and Crowdstorm: The future of innovation, ideas, and problem solving, by Shaun Abrahamson, Peter Ryder, and Bastian Unterberg (2013). In addition, the researcher reviewed both relevant studies and similar crowdsourced usability testing websites, such as Crowdsourcedtesting (https://crowdsourcedtesting.com/), Usabilla (https://usabilla.com/), and Loop11 (https://www.loop11.com/). Table 1 and Table 2 present the framework for the design of the proposed crowdsourced usability-testing tool based on Hanna’s guidelines and the cognitive load theory (CLT) (See Table 1 in page 810). Changes have been made later in the study through the iterations.

Table 2

Designing the Remote Usability Testing Tool Based on the CLT

<table>
<thead>
<tr>
<th>CLT Learning Effects</th>
<th>Tool’s Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>The worked-example effect</td>
<td>As a worked example, the tool in the current study provides a video of a child performing some tasks in a usability test.</td>
</tr>
<tr>
<td>The split-attention effect</td>
<td>All pieces of information that are needed to learn a specific knowledge component are integrated in one source. For example, instructions that are related to one aspect of the usability test are integrated in one video, screen, and/or device.</td>
</tr>
<tr>
<td>Effect</td>
<td>Description</td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>The modality effect</td>
<td>Visual information is integrated with necessary, non-redundant auditory (spoken) information where possible. This is shown in the videos that are provided as a website introduction and as a worked example of usability testing as well as all the other videos that were created throughout the iterations of the current study.</td>
</tr>
<tr>
<td>The redundancy effect</td>
<td>Redundant information is not integrated in one source.</td>
</tr>
<tr>
<td>The expertise reversal effect</td>
<td>Necessary information for novice users are removed for more experienced users based on the users’ profile information in the website/tool.</td>
</tr>
<tr>
<td>The guidance fading effect</td>
<td>The guidance provided to the parents during the testing session is gradually removed based on the users’ profile information in the website/tool.</td>
</tr>
<tr>
<td>The imagination effect</td>
<td>N/A</td>
</tr>
<tr>
<td>The element interactivity effect</td>
<td>The previous effects are especially relevant and need to be considered when there is high element interactivity.</td>
</tr>
<tr>
<td>The isolated interacting elements effect</td>
<td>Interacting elements are presented individually before highlighting their interaction. For example, as a test moderator, parents need to learn how to</td>
</tr>
</tbody>
</table>
Table 2 continued

moderate a usability test. In order to do so, they need to learn the different components that are part of the process. These components should be presented individually before explaining how they integrate to form the process of moderating a usability test.

**Development and pilot study.** A prototype of the crowdsourced usability testing website was developed based on the review of relevant studies, the review of similar websites, and based on both the cognitive load theory and Hanna’s guidelines (see Appendix A and B for storyboard and screenshots of the website). In general, crowdsourcing websites are structured to serve two types of users: The crowdsourcers and the crowd, or, in the current study, the instructional designers and the usability testers. Although this study focused only on the usability testers’ part of the crowdsourced usability test, the prototype’s user interface gives the test users the feeling that they are actually using a fully developed crowdsourcing website with both the instructional designers’ and the testers’ parts. This helps the parents understand the nature of crowdsourced usability testing websites, and it helps the researcher gain more relevant answers when interviewing the parents after using the website.

As shown in Figure 1 below, the general structure of the tool consists of four sections derived from Hanna’s guidelines: Set-up and planning, introductions, during the test, and finishing up. Under the set-up and planning section, the parent receives
instructions on how to set up the testing area as described in Table 1. The parent also completes a checklist to ensure that he or she did not miss any of the points and to inform the researcher of any unachievable arrangements. Then, the parent watches an introductory video that explains the goal of the usability test, provides a website tour, and provides instructions on how to moderate the test and report the results.

In the introductions section, the child completes an assent form, then he or she performs a warm-up activity to practice using the input device before starting the actual test. The warm-up activity includes using the touchpad to move the cursor, click, and drag objects on the screen. The activity was created by the researcher using Adobe Animate and is embedded in the child’s interface to ensure that the child does not face any inconvenient adds or navigate outside of the prototype. When the child completes the warm-up activity, an introductory video is played, which explains what will happen and what is the child’s role in the usability test.

When the child and the parent are both ready to start the session, they are instructed to operate the recording device/tool to capture the screen and to record the verbal data and the webcam. Silverback software is used to record the screen, sound, and webcam. During the test, the parent has access to a test moderator/parent page from which he or she reads the tasks for the child, types notes, answers task-specific questions while observing the child, and finds support and hints. At the completion of the testing session, the parent is prompted to review the report and to show gratitude for the child. Both the child and the parent receive gift cards as a gratuity for their time.
The website was reviewed by an elementary school teacher to ensure its suitability for children in terms of the language and the fonts that were used in the website. The teacher was selected randomly from a list that includes all third and fourth grade teachers who serve in Athens city schools, Ohio. The evaluation request was sent to the teacher using the email address provided in the Athens City School District online directory. The teacher “found it very user friendly, appropriate, easy to understand, and easy to follow. The font seemed appropriate, and there were no terms or tasks that seemed beyond a 4th grader's ability to understand” (personal communication, November 18, 2017). The teacher has identified other issues that are not related to the font and terms used but that could cause confusion to the children (Appendix C). All issues were related to the warm-up activity and the multimedia used in the children’s interface, and they were resolved based on the teacher’s recommendations. For example, the problem

Figure 1. The general structure of children’s crowdsourced usability testing
with the audio in one of the videos was fixed, and the narrative in the video was also modified to provide more specific instructions. Eight other teachers were sent the same email, each with an interval of two days, but none of them responded. At this point, the researcher decided to proceed with one teacher due to the time constraints.

The multimedia used in the website was reviewed by Fred Paas, professor and chair of the educational and developmental psychology program at Erasmus University Rotterdam, Netherlands (Appendix D). Professor Paas is known for his substantial contributions to the cognitive load theory. His research interests include applying the cognitive load theory to the design of instructions, multimedia, and learning environments and measuring cognitive load. He shares several publications with professor John Sweller, the developer of the theory. His email indicated that the multimedia used in the prototype does not violate the cognitive load theory, and that it seems appropriate for the intended age group of the children and their parents (personal communication, December 5, 2017).

After the necessary changes were applied, a pilot study was conducted which included two sessions, each with a nine-year-old child and her parent. The parents were both graduate students. Each session lasted about two hours. The children’s participation took around 30 minutes; the rest of the session was with the parent. The purpose of the pilot study was to examine the observation techniques and the quality of the audio and visual data. It was also conducted to ensure the appropriateness of the test set-up, the clarity of the tasks, and the minimization of the negative effects on children users (Markopoulos et al., 2008). Parents and children received $10 each at the end of the testing session to show appreciation for their time.
Some changes were applied to the website based on the issues that were identified in the pilot test sessions and based on the participants’ suggestions. These changes include adding an item to the checklist which prompts the participants to avoid sitting in front of a light source during the session in order to not ruin the quality of the recorded video; adding a note that prompts the parent to speak loudly right before starting recording the session; reminding the parent in the report page to count the time once the child starts the activity; adding a note in the report page that asks the parent to show the child the multiple choices before asking him or her to answer the question; and changing the title of one of the pages from “Starting the Usability Test” to “Starting the Session” since one of the participants indicated that the title was confusing because this was the first time she heard or read this term in the website.

After analyzing the data of the pilot sessions, the researcher decided to include the two sessions in the first iteration for two reasons. The first reason is that the changes that were required after conducting the pilot sessions were not major and were not expected to significantly affect the collected data. The second reason is that the nature of the current study is iterative; meaning that there were supposed to be changes in the prototype anyway after every third session, and it was difficult to consider some changes as pilot changes rather than considering them as an important part of the study, especially that the main topic of the study is usability testing and cognitive load, which overlaps with the objectives of the pilot study.

**Evaluation.** The developed instance needs to be tested in order to demonstrate two important points: The applicability of the instance in the proposed context and the viability of the results in solving the targeted problem (Ellis & Levi, 2010). Richey et al.
(2004) suggested the use of multiple resources of evidence, such as observations, interviews, questionnaires, and document analysis. In the current study, both observations, fieldnotes, and interviews were used to evaluate the proposed prototype and to indicate how it meets and does not meet the desired functionalities. An IRB approval was obtained before the data collection was initiated (Appendix E).

This process was conducted through three iterations, each with three evaluation sessions. This means that there were nine sessions in total, each included a child and a parent. New participants were recruited for each evaluation session. The sessions were conducted at the participants’ homes in order to test the product in its natural context. As discussed in the previous chapter, usability testing that is conducted in the field was found to provide higher ecological validity since it simulates the actual use of the product and facilitates the detection of real usability problems (Markopoulou et al., 2008). Each session included the testing of Tynker platform (https://www.tynker.com) by a child in the presence of his or her parent through the prototype website created for the current study. The parent’s role was to moderate the usability test and to submit the report that includes the outcomes of the test. Tynker is an educational platform that teaches coding to children seven years old and above through online games and courses. It meets certain standards of online safety and privacy for children as it has been reviewed and certified by kidSAFE® Seal Program.

Silverback software was used to record the child’s interaction with the website and with the parent by recording the screen, webcam, and sound on the child’s device. In addition, the researcher was observing the session and taking notes in order to answer the following research questions (see Appendix F for observation guide):
1. What can be learned about the remote usability testing guidelines in terms of:
   
a. the features and characteristics of the usability testing tool that are related to the children nine to ten years old?

b. improving the role of parents in such a way that can reduce their workload while maintaining a good quality of usability testing results?

Although the researcher was present, she was out of the participants’ sight as much as possible and did not intervene in the usability test or interact with the child or the parent during the session.

At the end of the testing session, the researcher interviewed the parent in order to answer the previously mentioned research questions in addition to the following question (see Appendix G for the interview protocol): What concerns do parents have regarding their children’s participation in crowdsourced usability testing?

Typically, at the end of this phase, the researchers decide whether to iterate back to the design and development phase or to move to the next phase leaving “further improvements to subsequent projects” (Peffer, Tuunanen, Rothenberger, & Chatterjee, 2007, p. 56). Due to the limited time and the limited number of participants who were willing to participate, the researcher decided to stop at the third iteration. At the end of this phase, the researcher was able to identify what worked well and what did not in the design of the tool as will be shown in the next chapter.

**Participants**

According to Richey and Klein (2014b), a small number of participants are usually recruited in qualitative design and development research, but it is difficult to tell the exact number of participants for such studies. Researchers usually start the process of
recruiting participants and collecting data “until no new information is discovered” (Richey and Klein, 2014b, p. 89). However, due to the limited time and number of participants who were willing to participate, especially that the study required videotaping the child and visiting the participant’s homes, the current study stopped at the third iteration. As discussed in Chapter 2, Nielsen (1994) suggests that three to five participants per iteration is enough to obtain sufficient results for a cost-effective usability testing. The participants of the current study included three children and three parents per iteration. Since this study consisted of three iterations, each with three parents and three children, there were eighteen participants in total, nine children participants and nine parent participants. The pilot test included two children with their parents.

The parents included eight females and one male from five different countries. They were between 33 and 48 years old. The children, who were eight females and one male, were nine to ten years old and have at least six months of experience with computers. This age range was selected for several reasons. According to Markopoulos et al. (2008), the think aloud method, which is usually used in usability testing, can be used with children as young as seven years old. In addition, children at this age have experienced the use of computers at schools for more than six months, which is the minimum requirement for usability testers in terms of computer experience according to Hanna et al. (1997). It is worth mentioning that the current study was conducted in a rural college town with a school district that has wireless internet and Chromebooks available for every grade in the entire district, which facilitated the sampling of children with at least six months of computer experience. Finally, the group was limited to
children of nine to ten years old to reduce the variation in the participants’ attributes. It is assumed that “there may be just as big a difference between second and third graders as there is between first and fifth graders” (Byerly, 2007, p. 32). “In general two major turning points can be assumed with key adjustments in methods being made for respondents older or younger than 7-8 years, and older or younger than 11-12 years,” which reflects Piaget’s stages of cognitive development (Lobe, Livingstone, Olafsson, & Simões, 2008, p. 15). Therefore, the age group of the participants in the current study was picked between these two turning points because children at this age are supposed to have sufficient computer experience and the ability to think aloud, and there are minimum differences in their attributes compared to other age groups.

According to Lobe et al. (2008), since the goal of qualitative research is different than the goal of quantitative research, sampling practices are consequently different for both types. In qualitative research, which does not aim to generalize the findings to a total population (Creswell, 2014), samples do not have to be representative of the population (Lobe et al., 2008). Researchers in this case decide which children’s characteristics are important in the sample, and then select the children with those characteristics. The sampling can be done in cooperation with the school, with parents, or through the researcher’s own social network (Lobe et al., 2008). In the current research, the researcher’s own social network was used to recruit the participants purposefully based on different demographics. The participants were contacted through text messages to participate in the study. They were also asked if they know of any other parents of nine to ten years old who could participate in the study and who have at least six months of computer experience. Table 3 shows the details of the participants in each
session. None of the participants had usability testing experience. The English language level of the participants was self-reported based on the following categories:

1. Basic: The participant can say and understand a few things in English.
2. Intermediate: The participant can use English in basic everyday context.
3. Advanced: The participant speaks and understands English with the occasional mistake.
4. Fluent: The participant uses English with ease and fluency.
5. Native: English is the participant’s first language.

Table 3

Participants’ Details

<table>
<thead>
<tr>
<th>Participant</th>
<th>Gender</th>
<th>Age</th>
<th>English Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent 1</td>
<td>Female</td>
<td>30-34</td>
<td>Fluent</td>
</tr>
<tr>
<td>Child 1</td>
<td>Female</td>
<td>9</td>
<td>Fluent</td>
</tr>
<tr>
<td>Parent 2</td>
<td>Female</td>
<td>35-39</td>
<td>Advanced</td>
</tr>
<tr>
<td>Child 2</td>
<td>Female</td>
<td>9</td>
<td>Advanced</td>
</tr>
<tr>
<td>Parent 3</td>
<td>Male</td>
<td>40-44</td>
<td>Fluent</td>
</tr>
<tr>
<td>Child 3</td>
<td>Female</td>
<td>9</td>
<td>Native</td>
</tr>
<tr>
<td>Parent 4</td>
<td>Female</td>
<td>35-39</td>
<td>Fluent</td>
</tr>
<tr>
<td>Child 4</td>
<td>Female</td>
<td>9</td>
<td>Native</td>
</tr>
</tbody>
</table>
Table 3 continued

<table>
<thead>
<tr>
<th>Parent 5</th>
<th>Female</th>
<th>35-39</th>
<th>Fluent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child 5</td>
<td>Female</td>
<td>9</td>
<td>Native</td>
</tr>
<tr>
<td>Parent 6</td>
<td>Female</td>
<td>35-39</td>
<td>Advanced</td>
</tr>
<tr>
<td>Child 6</td>
<td>Female</td>
<td>10</td>
<td>Native</td>
</tr>
<tr>
<td>Parent 7</td>
<td>Female</td>
<td>30-34</td>
<td>Native</td>
</tr>
<tr>
<td>Child 7</td>
<td>Female</td>
<td>9</td>
<td>Native</td>
</tr>
<tr>
<td>Parent 8</td>
<td>Female</td>
<td>45-49</td>
<td>Fluent</td>
</tr>
<tr>
<td>Child 8</td>
<td>Female</td>
<td>10</td>
<td>Advanced</td>
</tr>
<tr>
<td>Parent 9</td>
<td>Female</td>
<td>35-39</td>
<td>Native</td>
</tr>
<tr>
<td>Child 9</td>
<td>Male</td>
<td>9</td>
<td>Native</td>
</tr>
</tbody>
</table>

According to Richey and Klein (2005), each phase in the design and development research usually has different types of participants, which is the case in the current study. There are two different types of participants in addition to the children and the parents. The first is an elementary school teacher, who was selected randomly from a list that includes all third and fourth grade teachers who serve in Athens city schools, Ohio. The teacher reviewed the tool to ensure its suitability for the age of participants. The second is Fred Paas, who ensured the tool’s correspondence to the principles of cognitive load theory. Fred Paas is professor and chair of the educational and developmental
psychology program at Erasmus University Rotterdam, Netherlands, and he is known for his significant contributions to the cognitive load field. His research interests include measuring cognitive load and applying the cognitive load theory to the design of instructions, multimedia, and learning environments.

**Data Collection and Analysis**

In design and development research, it is critical to apply multiple data collection methods, especially when the researcher is the designer of the developed product, which invites the possibility of flaws in data credibility and integrity (Richey & Klein, 2014b). Therefore, in this type of research, multiple data collection methods are usually used to triangulate the data and reduce personal bias (Richey & Klein, 2014b). There are four basic types of qualitative data collection methods: Observations, interviews, document analyses, and audio and visual materials analyses (Creswell, 2014). In the current study, three data collection methods were used to answer the research questions: Observations, fieldnotes, and interviews.

The purpose of observations is to discover complex interactions and behaviors as they occur in their natural settings (Creswell, 2014; Marshall & Rossman, 2011). They entail “the systematic recording of events, behaviors, and artifact (objects) in the social setting” (Marshall & Rossman, 2011, Chapter 6, Section 1, para. 1). There are different types of observations that vary in the level of the observer’s participation in the social event. According to Creswell (2014), the researcher has four options: Complete observer (the researcher observes without participation), participant as observer (the researcher is more a participant than an observer), observer as participant (the participants know the role of the researcher), and complete participant (researcher’s role is unknown to the
participants). In the present study, the researcher role was of the observer as participant type; the researcher was a passive participant.

Fieldnotes are necessary to record what has been observed during the session (Marshall & Rossman, 2011). Researchers often use observation guides, such as observation checklists and forms, which direct them to notice specific behaviors during the observation (Richey & Klein, 2014b). In early stages of the research, though, researchers usually approach the observation broadly without predetermined checklists. Checklists and the predetermined categories are created after analyzing the first fieldnotes and discovering the repeated patterns of behaviors, which allows for more focused observation at the later stages of the research (Marshall & Rossman, 2014). In the current study, the researcher used an observation guide and recorded the fieldnotes. Observation checklist were then created after analyzing the first fieldnotes.

Since it is difficult to interpret the observed interactions (Marshall & Rossman, 2014), researchers often follow up observations with interviews (Richey & Klein, 2014b). The purpose of interviews is to “get a clear understanding of events, to determine why they occurred, and to gather data from participants about their thoughts and beliefs” (Richey & Klein, 2014b, p. 113). In other words, interviews help the researcher connect what the participants do during the observation with what they say in the interview in order to achieve a better understanding of their behaviors. Design and development researchers use interview protocols to guide them through interviews and to ensure that all interviewees receive the same prompts (Richey & Klein, 2014b). Patton (2002) categorized interviews into three types: the informal conversational interview, the interview guide approach, and the standardized open-ended interview. The three types
vary in the level of adaptability that they allow during the interview. The interview guide approach was adopted in the present study since it is semi-structured, which allows the researcher to have a moderate level of flexibility and openness while being prepared with a set of flexible questions before the interview.

**Data collection procedure.** The study was conducted at the participants’ homes where the researcher met the parent on a previously scheduled time. The parent completed a consent form that explains the purpose of the study; what will the participants be asked to do; how and why the sessions will be recorded; and the freedom to opt out of the study at any time without a penalty (Appendix H). The researcher emphasized that what is being tested is the product itself and not the child, as recommended by Nielsen (1993b) and Markopoulos et al., (2008), to minimize the bias of the parents who might think that their children are being tested. A description of the product and assurance that it will cause no harm was also provided.

After completing the consent form, the parent accessed the usability testing website through the parent’s device and interface in which he or she read a list of set-up instructions and completed a checklist to ensure that all the important points in the instructions are achievable and to inform the researcher of any circumstances that interfere with the desired context and might affect the test. Next, the parent watched a video that presents the purpose of the website and that explains the interface and the role of the parent. When the parent completed the set-up and the tour and instructions sections, he or she was given a code so the child can access the children’s interface and complete the assent form, which has the same elements as the consent form only with a simple language that is easier to understand (Appendix I). The assent form also
emphasizes the importance of the child’s role in making better products for other children. Once the child accepted to participate in the study, he or she started a warm-up activity through the children’s interface to practice using the input device before beginning the usability test. The researcher’s involvement was minimized as much as possible throughout the session.

When both the parent and the child were ready to begin testing the product, the child watched a video that presents the purpose of the test and shows an example of a person performing a usability test. At the end of the video, the child was encouraged to ask the parent any questions regarding the test. When ready, the parent was given instructions on how to start recording the session, and the child was taken to Tynker, the website that needs to be tested. The recording tool recorded the screen, the audio, and the webcam. The parent on the other hand was accessing the other interface in which he or she read the tasks for the child one at a time and entered relevant notes regarding the usability of the website for each task that the child is working on. The parent was prompted to encourage the child to verbalize his or her thoughts through a dialogical approach in order to reduce the cognitive load that might be otherwise increased if the original approach of thinking aloud was used (Markopoulos et al., 2008).

At the completion of each task, the parent answered some questions to rate how difficult it seemed to be for the child to complete the task from the parent’s observation. At the end of the tasks, the parent asked the child a number of questions using open-ended questions and likert scale questions to rate the usability of the website. Using different techniques allowed the researcher to record the child’s opinions, thoughts, and emotions in different ways that do not rely heavily on a single ability, such as
verbalization (Khanum & Trivedi, 2012). The parent’s interface has a support page that provides tips for dealing with different situations in the test. The parent was able to access the support page any time during the test. The researcher was observing and taking notes throughout the session without interfering unless it became impossible to complete the session without help. At the completion of the tasks, a note at the parent’s interface asked the parent to review the usability report, adding any important observations, and the child received ten dollars to show gratitude.

Afterward, the parent was interviewed to obtain an insight on how to improve the parent’s role in a crowdsourced usability testing website in such a way that can reduce the workload while maintaining a good quality of usability testing results. The interview also helped in understanding the concerns that the parent has regarding the child’s participation in crowdsourced usability testing in terms of safety, time, and reward. The researcher used the interview guide approach (Patton, 2002); she followed a set of outlined topics with some degree of flexibility in the wording, order of questions, and probing. Member checks were applied during the interview to improve the validity and credibility of the study (Cohen, Manion, & Morrison, 2011; Cresewell, 2014); the researcher summarized the parent’s answers and asked the parent to determine the accuracy of the information right after each question during the interview. An audio recorder was used to record the interview, which was transcribed, and the fieldnotes were expanded as soon as possible after leaving the field site. Follow-up questions were emailed to the parent when needed.
Figure 2. Data collection procedure
**Data analysis.** Data analysis is “the process of making sense out of the data” (Merriam & Tisdell, 2016, p. 202). Unlike the data analysis in quantitative research, the analysis of qualitative data proceeds “hand-in-hand with other parts of developing the qualitative study, namely, the data collection and the write-up of findings” (Creswell, 2014, p. 195). Merriam and Tisdell (2016) explained that ongoing analysis prevents the data from being “unfocused, repetitious, and overwhelming” because the researcher in this type of research knows the problem but does not know “what will be discovered, what or whom to concentrate on, or what the final analysis will be like” (p. 197).

In the present study, data was collected using observations and interviews, and it was analyzed following the six steps suggested by Creswell (2014):

1. Organizing and preparing the data.
2. Reading through all the data.
3. Coding the data.
4. Generating descriptions and themes.
5. Interrelating themes and descriptions.
6. Interpreting the meaning of themes and descriptions.

To store and code the data, the researcher used MAXqda qualitative analysis software. Coding is the “process of making notations next to bits of data that strike you as potentially relevant for answering your research questions” (Merriam & Tisdell, 2016, p. 204). The researcher decided to use qualitative data analysis software because hand coding is more time-consuming and less efficient (Creswell, 2014).
The Role of the Researcher

Since the researcher is the key instrument in qualitative research, the researcher’s background, experience, and attributes can influence the direction of the research findings (Creswell, 2014; Patton 2002). Therefore, these aspects should be explained in order to acknowledge the possible biases that might reshape the interpretations and findings. This reflection should enhance the trustworthiness of the research (Marshall & Rossman, 2011). In this section, the background, the role, and the potential biases of the researcher will be explained.

The researcher’s interest in researching usability testing resulted from her major in programming and instructional technology and from her job as a university lecturer in the instructional technology field. The researcher not only realizes, but has also faced the challenges of usability testing. During her research, she realized that crowdsourcing has been used to overcome the challenges of usability testing lately, but it is not being used to test children’s products. As a mother of two children, she is aware of the value of finding high quality educational products and materials for children that are not a waste of time and money. For these reasons, and from her desire to improve the field, her curiosity in researching crowdsourced usability testing of children’s products has emerged.

The researcher’s understanding of the context due to her background and experience enhances her awareness and sensitivity to many of the issues encountered in usability testing as an instructional technologist and as a parent. However, the researcher’s background and experience can be a source of bias. Furthermore, the fact that the researcher is also the designer of the tool under investigation might hinder the validity of the research (Richey & Klein, 2005; Richey & Klein, 2014b). Therefore, it is
important to follow certain strategies to enhance the validity of the research and reduce potential bias (Creswell, 2014). The data in this research will be triangulated through interviews and observations in order to “build a coherent justification for themes” (Creswell, 2014, Chapter 9, Section 9, para. 4). Member checking will also be used to enhance the validity of the participant’s responds. Thick description of the sessions will also be provided in order to present more realistic and richer results to the reader (Creswell, 2014).

The researcher’s role in the current study was observer as participant (Creswell, 2014). More specifically, the researcher was a bystander; she limited her interaction with the participants during the observation session. This is due to the difficulty of being a complete observer in the settings chosen for the study. The participants’ homes are not prepared for complete observations where the participants are not aware of the role of the researcher who is hidden from view. The researcher chose to minimize her participation in the study because the goal of this research is to investigate the use of crowdsourced usability testing at home instead of the lab, where test administrators interact with and observe the participants; being a participant observer means hindering the goal of the research.

**Methodological Issues**

Richey and Klein have determined several critical methodological issues related to design and development research (Richey & Klein, 2005; Richey & Klein, 2014b). The first issue is “the need to account for contextual variables” (Richey & Klein, 2005, p. 33). They clarified that since it is impossible to control for these types of variables, the researcher needs to “carefully describe (and measure, if possible) these variables in an
effort to account for their impact” (Richey & Klein, 2005, p. 33). In the current study, a rich description of the context was provided. This includes the region, the gender of the participants, the education level of the parent, type of technology used during the test, a description of the physical layout of the space in which the test is conducted, and the English language proficiency of both the child and his or her parent.

Although the lessons learned from design and development studies of products and tools can be applied to other similar studies, they tend to be less generalizable than model studies since they are generally context-specific (Klien, 2014). Data from a specific project can be useful for a broader variety of projects in the workplace if it is illustrative of those projects (Ellis & Levi, 2010; Richey & Klein, 2014b). Qualitative researchers, however, argue that “case-study data, such as those describing single projects, can often provide explanations superior to those based upon data from a broad population of participants” as they allow the researcher to explore the topic more intensely (Richey & Klein, 2014b, p. 140). Nevertheless, generalizations cannot be made before conducting a series of design and development studies in order to reach an acceptable level of certainty (Richey & Klein, 2014b). Therefore, suggestions proposed in the current study cannot be generalized until sufficient replications are conducted.

Another issue is that in this type of research the researcher is also the designer and the developer of the tool, which might hinder the validity or the credibility of the research (Richey & Klein, 2005; Richey & Klein, 2014b). They suggest that in these cases “care must be taken to insure objectivity through consistent, systematic data collection techniques and the collection of corroborating data, if possible” (Richey & Klein, 2005, p. 33). In order to ensure the validity and credibility of the current study, the accuracy of
the findings was gauged through triangulation and member checks (Cohen, Manion, & Morrison, 2011; Creswell, 2014). The participants were observed during the test and interviewed after the completion of the test. As soon as the data reports were completed, they were reviewed by the participants in order to increase the accuracy and the credibility.

Another issue that can also be solved with data triangulation is related to data integrity, especially when interviews and self-reports are used in the study (Richey & Klein, 2005). It is common for researchers to use self-report data as it has several advantages. Self-reports offer data directly from the respondents, they provide rich data, and they are easier to conduct than other data collection methods (Dodd-McCue & Tartaglia, 2010). The accuracy of self-reports data is questioned though as there are several factors that could bias the data (Richey & Klein, 2014b). Some of these factors are social desirability, acquiescence, leniency and harshness, and extreme and midpoint response style (Dodd-McCue & Tartaglia, 2010). Collecting data from multiple resources enables the validation of data and minimizes the bias (Richey & Klein, 2014b). However, sometimes it is hard to avoid the bias in self-reported data, and in these cases, it is crucial to acknowledge that as a limitation in the study.

An issue that is more relevant for product and tool researchers is “recognizing the influence of work environment characteristics” (Richey & Klein, 2014b, p. 61). Product and tool researchers tend to collect data in real world settings in order to increase the authenticity of the research. However, they need to recognize the influence of the contextual factors that might affect the processes and the outcomes of the research. Researchers can recognize the influence of contextual factors by providing thick
description of the environment and the settings of the study as well as the people involved in the design of the product or tool (Richey & Klein, 2014b).

**Ethical Consideration**

The current study was conducted only after receiving the Institutional Review Board (IRB) approval. According to Markopoulos et al. (2008), usability testing has to be carefully planned and implemented in order to protect young and adult participants from physical and psychological harm. In order to reduce the chances of risk, a pilot study was conducted. Informed consents and assents were collected from the parents and the children. In order to reduce the pressure and anxiety in both the children and their parents, it was emphasized that the product is what is being tested and not the child (Nielsen, 1993b). It was also explained to the children that they can opt out of the study at any time without consequences. Any noticed desire to opt out in the child’s body language or behaviors was taken into consideration and the child was asked if he or she wants to withdraw. It was decided that in case a child wanted to withdraw, he or she would still receive the promised incentive.

Nielsen (1993b) assured that the privacy of the participants’ identities and the confidentiality of the test results have to be maintained. The records of the current study will remain private, and the data was aggregated. Direct identifiers like the participants’ names were replaced with codes in the stored records. Any information that can make it possible to identify the participants will not be disclosed in the public report without the participant’s permission; pseudonyms will be used to ensure anonymity if needed (Ritchey & Klein, 2014). Both the parent and the child were informed that they were going to be video and audio recorded during the study. Recordings were not shared
outside the research team and will be destroyed approximately three years from the completion of the research.

Since the test’s tasks involve using web content, two issues were taken into consideration (Markopoulos et al., 2008); protecting the child from being exposed to inappropriate content and from sharing personal information. For example, the researcher programmed the warm-up activity for the child instead of using an online activity that might include inappropriate adds or that might take the child to an external website. Special attention was given to the difficulty and the length of each task in order to avoid distressing the child (Markopoulos et al., 2008). Clear instructions were provided to the parent as a test administrator in order to protect the child from being distressed during the test. In case the child became distressed for any reason, the researcher would have stopped the session immediately (Markopoulos et al., 2008).

**Summary**

In this chapter, the researcher presented the design and development methodology that was used in the current study. Other details such as the research participants and the data collection procedures were provided, and the methodological and ethical concerns were discussed. The next chapter presents the findings of the study.
Chapter 4: Findings

Introduction

In this study, the researcher designed and developed a prototype of an online usability testing tool in order to create guidelines for remote usability testing with children. The goal of the tool is to connect instructional designers with children in order to facilitate the recruitment of usability testers for children’s learning products. The prototype was designed based on Hanna’s et al. (1997) guidelines and the cognitive load theory. Eighteen participants were recruited to try the prototype and help improve the guidelines; the children in the study are referred to as ‘child n,’ and the parents are referred to as ‘parent n,’ whereas ‘n’ in both cases is the session number. Data was collected through three iterations using three methods: Observations, fieldnotes, and interviews. After each iteration, changes were applied to the prototype based on the data analysis (see Appendix J for the changes applied in each iteration).

In this chapter, the findings that are related to the following research questions will be provided:

1. What can be learned about the remote usability testing guidelines in terms of:
   a. the features and characteristics of the usability testing tool that are related to the children nine to ten years old?
   b. improving the role of parents in such a way that can reduce the workload of parents while maintaining a good quality of usability testing results?

2. What concerns do parents have regarding their children’s participation in crowdsourced usability testing?
Children-Related Features

What happened? There are three stages that the children went through the first time they used the usability testing website developed in this study. The first is the introduction stage, which includes the assent form, a warm-up activity, and the animated introduction video. Before starting the session, the children read the assent form provided in the children’s interface of the website (see Appendix B and Appendix I). After they read the assent form and accepted to participate in the usability test, the children went through an ad-free warm-up interactive activity that ensures that the child does not start the session without being able to use the input devices required to try the product to be tested. In order to complete the warm-up activity, children had to successfully use the basic touchpad skills: Clicking and dragging. Then, they watched an introduction video that includes the goal of the session, an explanation of what will happen in the session, and a short example of a child performing a usability test. In the next stage, the children went to the product to be tested and started performing the tasks read by the parent. At the end of the last task, the children answered some questions related to the usability of the product. The last stage is the finishing-up stage, in which the children are given a reward as a gratitude for their time and help. In this study, the children’s reward was a ten dollars gift card.

Overview of the children-related findings. This section includes a summary of the findings that are related to children. The next section includes a detailed description of these findings.

1. Some children found difficulties reading the font that was appropriate for most children.
2. Children showed positive reactions when they were told that their parents are going to be with them during the session.

3. Children seemed comfortable while they were being video recorded during the session, in which their faces were not showing on the screen.

4. During the session, children imitated what they saw in the video example of a usability testing session.

5. Children imitated their parents during the session in talking or staying silent, speaking loudly or softly, and in speaking in a specific language.

6. Some children were shy of telling their sincere opinion about the product.

7. When they were asked about their opinion in the website they tested, children were confused between that website and the usability testing website.

8. Children seemed bored when parents were slow in typing their feedback.

9. When parents discussed the answers with the children, the children’s answers were more detailed than when they entered their answers without parents’ presence or interference.

10. Children did not seem tired, and they did not ask for a break during the sessions, which lasted between 37 and 63 minutes.

11. Some children seemed anxious about the timer.

12. Children were excited and happy to earn the reward at the end of the session.

**What went well and what didn’t?** This section is organized based on the stages that were described above: the introduction stage, during-the-test stage, and finishing-up stage.
**Introduction stage.** In this stage, the children read the assent form, performed the warm-up activity, and watched the instructions video.

**Assent form.** Although the assent form was reviewed by an elementary school teacher, who confirmed that it was suitable for children nine to ten years old in terms of the language and the font size, and although the children in eight of the sessions did not seem to have a problem, there was one child who seemed to have difficulty reading the text. She brought the laptop’s screen closer to her face while reading. Her parent later suggested:

The part that [my daughter] was supposed to read, the letters were small, but then thinking from like an accessibility standpoint, maybe someone who wears glasses or have some vision problems, maybe if there’s the option to scroll that thing and increase the font size, or maybe just the font size is a little bigger.

No other observations were noticed regarding the assent form. All the children read the assent form by themselves and did not have any questions.

**Warm-up activity.** By accepting to participate in the usability test, the children were taken to the interactive warm-up activity, which allows the child to practice clicking and dragging using the laptop’s touchpad. These skills were required to be mastered in order for the children to be able to use Tynker, the product to be tested. The warm-up activity was designed following the cognitive load theory principles discussed in Chapter 2, and it was narrated by a child. None of the children showed any difficulties in completing the activity. They all completed it in three to four minutes. It is worth mentioning that the profiles the parents created indicate that all of the participating children have been using laptops among other devices. When the parents were asked if
there were any unnecessary parts in the experience they had, parent 4 mentioned the
warm-up activity. She said:

In the section for the children, which teaches them how to drag, I feel like even
though I know the kids are familiar with it, I feel like it’s important to get them
engaged, because someone is speaking with their voice, and it’s very nice, even if
it’s like long and I feel like not necessary, but I feel it’s important to engage the
kid.

There were no other observations or comments about the warm-up activity.

*Introduction and instructions video.* When the children completed the warm-up
activity, they moved to the introduction and instructions animated video. In the first
iteration, which included the first three sessions, this video was about five minutes long,
and it included the goal of the session, what will happen in the session, and a short
example of a child performing a usability test with her parent. The example showed the
child’s screen and the voices of the child and the parent. The internet connection in
session 1 and session 4 was slow, which affected the video streaming. The videos were
pausing constantly. Child 1 and parent 1 seemed slightly bored after the constant
pausing. Child 2 and child 3, who had no problems with the internet connection, were
smiling while watching the video. Child 2 started thinking aloud with the child in the
example in response to the mother’s questions or tasks.

None of the children in the three sessions repeated any parts of the video or asked
any questions. They seemed engaged. However, when the testing session started, the
child 1 and child 2 were imitating the example that was shown in the video. For
example, the child in the instructions video, who was testing a website that provides math
games, said that she thinks that the website teaches math, and then she started trying some of the math games. Although Tynker was a website that provides coding games, when parent 2 asked child 2 what she thinks this website is about, the child said that she was not sure, and then she said that it is a math website, which is the same as the example shown in the instructions video. When the parent asked her why she thinks it is about math, the child said, “because it said that in the video.” Similarly, child 1 said, “are there any grades?” and started looking for a button that shows grade levels like in the video she watched even before her parent asked her to do that. Therefore, the researcher decided to remove the example from the video. The video in the second iteration was one minute and fifty seconds long, and it included only the first two parts of the video in the first iteration, which are the goal of the session, and what will happen during the session.

The second iteration included sessions 4 to 6. Again, none of the children repeated any part of the video, and none of them had any questions when the video ended. As mentioned earlier, the internet connection in session 4 was poor, which caused slow streaming although the video size was fourth the size of the video in the first iteration. What was noticeable in this iteration is that parent 5 and child 5 were whispering during the whole session, and that parent 6 and child 6 were confused and the parent did not read the tasks for the child. In general, the children in the first iteration were imitating the child in the example they watched in thinking aloud, but the children in the second iteration had less verbalization or were whispering. The researcher decided to return the example to the instructions video in the third iteration but with some changes. The website that was tested in the example was of a different scope and style than the website that is to be tested in the session. This change was applied to try to
prevent the children from imitating the example, and at the same time to try to give them
and their parents a feel of how a usability testing session should be. In other words, the
researcher wanted the participants to imitate reading the tasks and the thinking aloud, but
did not want them to say things only because they were mentioned in the video. The
example this time also included negative feedback about the website since it was noticed
in the first iteration that the children were almost only giving positive feedback. One
child seemed shy to say her true opinion on the game that she could hardly play because
of the usability problems she faced during the session. Therefore, the example in the
third iteration included negative feedback. For example, when the mother in the video
asked her daughter what she thinks about the website, the child said, “I think it’s boring.”
Including explicit negative feedback in the example was an attempt to show the children
that it is fine to give such feedback.

The third iteration included sessions 7 to 9. The instructions video, including the
new example of a usability test, was three minutes and eight seconds long. When the
child in the example said that the website was boring, child 7 and parent 7 looked at each
other and smiled. No negative imitation was observed in any of the three sessions. The
parents were speaking loud and clear most of the time, and the children provided both
positive and negative feedback.

An observation that was noticed several times across the sessions was the
children’s reaction when the narrator said, “your parent will stay with you to manage the
session and to help you whenever you need help.” The children seemed comfortable and
happy to hear that. For example, when child 7 heard that sentence, she looked at her
parent and hugged her arm with a smile on her face. Child 9 also smiled to his parent and seemed relieved. Some other children had positive reactions as well.

**During-the-test stage.** The findings in this stage can be divided into three main points: Recording the session, thinking aloud, and answering the report questions.

*Recording the session.* At the end of the video, the parents started recording the session on their children’s devices. The software used to record the session is Silverback. In this software, the child can see his or her face on the screen to make sure that the camera is working before clicking on the Begin the Session button. Once the button is clicked, the child cannot see himself or herself. When child 2 read in the assent form that her face and voice will be recorded, she laughed in surprise. The way she laughed seemed like she was shy of being recorded. The parent explained later in the interview that her child “does not like anyone to videotape her or take photos of her.” She also said that “she was thinking of it all the time [during the session]. She was thinking that someone was watching her, just because she doesn’t like photos and videos.” However, during the observation, the child did not seem like she was uncomfortable. At one point, she was humming a song, and in general, she seemed enjoying the experience.

Although most parents expressed their concerns regarding videotaping their children, as will be shown in the parents’ concerns section, most children seemed comfortable about it. Although child 1 seemed aware of the researcher’s presence and looked at her couple of times during the session, she seemed comfortable and excited. Children 2, 7, and 9 were either humming or singing a song at one point during the session. Child 8 looked at herself on the screen and said, “I want to look pretty” while fixing her hair. Child 9 smiled at the camera when he saw his face, and he made funny
sounds later while he was playing. Child 7 was moving excitedly right after the recording has started that her parent told her to not move the laptop so it does not affect the recording. Similarly, child 5 was dancing with her head one minute after the recording has started.

One of the recording problems noticed in the first iteration was that the child 2 was sitting behind a light source. Later when the researcher watched the recording, she found that the light has made it hard to see the child’s face. Therefore, the second iteration included one more step in the set-up instructions. The parents were instructed to not let their children sit behind a light source.

Thinking aloud. One of the instructions given to the children was to think aloud. Also, the parents were instructed to encourage their children to verbalize their thoughts through conversation in order to discover usability problems. As mentioned earlier, in the first iteration, the children watched an example of a child thinking aloud and answering her parent’s questions while testing a website. Child 1 imitated the child in the video in thinking aloud. She was telling what she is trying to do in every step, even though her parent was not always responding or having a conversation with her; most of the time, the parent only read the tasks for her child and took notes. In this session, the internet connection was slow, and it affected the streaming of the parent’s instructions video.

In the second and third sessions, when the tasks were more specific, the parents tried to let their children verbalize their thoughts and explain their answers. For example, for the task that asked the child to take a look at the homepage and say what the website is about, the parent in the second session had a conversation with her child to let her talk
more and to understand why it was hard for her to figure out what the website is for.

However, when the task was more open, like the last task that asked the child to start playing in one of the games for ten minutes, there were long silence periods. Neither the child nor the parent talked. Later in the interview, the parent said that it was hard for her to have a conversation because she did not understand what her child was doing, and because she does not know a lot about games. Also, when child 3 said, “I don’t know how to play this game,” the parent said, “you don’t know? Okay, you can try another one.” He did not encourage her to talk more about why the game was difficult to understand. It is worth mentioning that there was one common problem with all the parents in the first iteration; they did not start the session by reading the first task and typing what they notice and then reading the second task and so on. They were all confused at the beginning of the session. The second iteration included clearer instructions for the parents, as will be discussed later, to help them start the session properly and more smoothly.

In the second iteration, although the example of the child thinking aloud was removed from the instructions video, two of the children did verbalize their thoughts. Child 6 did not, and when she talked with her parent, it was in Korean, the parent’s first language, even though the child was a native English speaker. In general, the parent did not try to have a conversation with her child, and the child did not try to think aloud. It was noticeable in the three sessions that the children’s verbalization depended on the parents; the sessions where the parent encouraged the child to talk had more verbalization than the session where the parent did not talk. On the contrary, in the first iteration in which the children watched a video example of thinking aloud, child 1 did well in
thinking aloud even without her parent talking. Another observation was that although the parent 5 and child 5 were having a conversation, they were whispering, and it was hard to hear their voices, both during the observation and in the recording, even though there was a note for the parent to speak loudly. Again, child 5 adjusted her voice volume depending on the parent’s voice volume.

As mentioned earlier, the instructions video in the third iteration included an example of a child thinking aloud in a usability test. Also, the note in the instructions that asked the parent to speak loudly was emphasized by making it bold. In sessions 7 and 8, the parents did well in having a conversation with their children. When there were silence periods from child 7, parent 7 asked, “what are you looking at?” a couple of times. Also, she tried to ask more detailed questions that were not mentioned in the report in both the specific and the open-ended tasks. For example, when the child was looking at a list of backgrounds with some locked and unlocked backgrounds, the parent asked her child about the lock symbol, what it meant, and if she knows how to unlock it. These questions were not mentioned in the report, but the parent was trying to discover more usability problems. In this session, the child’s answers were long, and she seemed comfortable.

Similarly, parent 8 also encouraged her child to think aloud. She explained the meaning of thinking aloud to her child at the beginning of the session. At one point in the session, the child said, “it’s frustrating,” and the parent replied, trying to let her child explain the problem, “why? What’s frustrating?” Their conversation was mostly about figuring out how to play the game. When the child said, “I don’t get this,” the parent said, “do you remember what you used to do with your robots?” The child said, “I know,
but I don’t know where the blocks are.” The parent added this as a problem in the report, but she said later in the interview that her lack of knowledge about coding affected her performance during the session. She said:

If there was something I knew beforehand, then I’d probably be better at searching clues and giving her some, you know, tips to help her finding them herself, but since I was as lost as her, I was like “okay, what should I tell her now.”

She also said that this problem caused her to concentrate mostly on the more obvious usability problems like text that is too small or hard to read, for example. It is worth mentioning that although the parent and the child were both native Spanish speakers, and the child’s English language as described by the parent was not fluent but advanced, they only spoke in English during the session.

Session 9 was the opposite of session 8 because the parent was an IT specialist and an IT graduate student. She understood the coding game, but she did not interfere and stayed silent most of the time during the open-ended task in the session. Although her son, child 9, did not know how to play the game properly, she did not guide him or give him any hints. She was watching him and typing her observations. Consequently, the child did not talk either. After trying several times to let the game work, he said, “I think I’m probably done,” and the parent laughed. It seemed that she knew that he did not know how to play, but she did not say anything. However, at the beginning of the session when there were more specific questions or tasks, the parent tried to have a conversation with her child to understand what he is thinking of. For example, when she asked him what the website was about, the child said that it was a games’ website. She
said, “did you look at everything? All the titles at the top of the page?” The child said, “yes”. She said, “but you think the website is about games.” He said, “aha.” There were no such conversations during the open-ended task in which the child was supposed to pick a game and play for a few minutes. Later in the interview, parent 9 explained:

I wanted to help him more than I was allowed, but I also wanted to see what he was able to do on his own just by looking and doing things that I have already taught him to do, I wanted to see what he did so I can talk to him later.

Parent 9 thought that she was not allowed to help her child, which revealed a problem in the parents’ instructions that will be discussed later in the parents-related features’ section.

*Answering the report questions.* As mentioned earlier, the usability testing report has two parts. The first part includes the tasks that the parents are supposed to read for their children along with some questions for the parents to answer based on their observations. This part is completed while the child is using the product under investigation. The second part of the report is supposed to be answered by the children after they are done testing the product, and it includes some questions about the usability of the product being tested. This section will discuss the findings that are related to the second part of the report. It is worth mentioning that the variation in the children’s opinions about the product in this part of the report was due to the fact that each of them picked and tested a different activity or game in the product. The findings that are related to the first part of the report will be discussed in the parents’ related features’ section.

In the first iteration, the parents were supposed to ask the report questions and type or enter the children’s answers. However, in the third session of the first iteration,
parent 3 did not read the line at the beginning of the report page that tells him to read the questions for the child and type her answers. He gave child 3 his laptop and told her to answer the questions, but he also discussed the answers with her. He said later in the interview that he is the kind of person “who really doesn’t read instructions a lot.” He said, “I try to figure it out very quickly.” Therefore, the line that tells the parent to read the questions for the child was underlined in the second iteration. Another observation in the first iteration was that parents 1 and 2 did not let their children see the multiple choices in the multiple-choice questions; they only read the head of the questions and let their children answer it without knowing the available choices. Also, child 1 was still playing in Tynker while answering her parent’s questions. The answers of child 1 and child 2 in sessions 1 and 2, in which the parents typed their children’s answers, were either brief or written from the parent’s point of view. The answers of child 3, in which the child typed her own answers, were more detailed compared to the two other children. However, in the last question, which asked the child if she wants to add anything that can help improve the website, child 3 discussed with her parent her detailed answer about the directions being unclear, and how she preferred to hear the directions rather than reading them. When they were done discussing her answer, he told her “okay, type it very shortly,” so she typed “put directions,” and in a new line she typed “voice.” Later in the interview, parent 3 explained, “I wanted to make it easier for her so she doesn’t have to write everything.” Table 4 shows the answers of the three children.
<table>
<thead>
<tr>
<th>Question</th>
<th>Child 1’s Answers</th>
<th>Child 2’s Answers</th>
<th>Child 3’s Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>How would you rate Tynker?</td>
<td>Awesome</td>
<td>Really good</td>
<td>Good</td>
</tr>
<tr>
<td>How much did Tynker make coding fun?</td>
<td>As much as possible</td>
<td>Very much</td>
<td>Very much</td>
</tr>
<tr>
<td>How easy was it to use Tynker?</td>
<td>Quite easy</td>
<td>Quite easy</td>
<td>Okay</td>
</tr>
<tr>
<td>What do you like best about Tynker?</td>
<td>Everything</td>
<td>That there is Minecraft. Play.</td>
<td>I like the candy quest game and I also I like how it teaches coding.</td>
</tr>
<tr>
<td>What do you like least about Tynker?</td>
<td>Nothing really</td>
<td>She felt somebody was watching her</td>
<td>I don't like the game brick breaker because they did not put directions and some of the blocks didn't fit.</td>
</tr>
</tbody>
</table>
Table 4 continued

<table>
<thead>
<tr>
<th>How easy was it to understand the activity you picked?</th>
<th>Really easy</th>
<th>Quite easy</th>
<th>Quite easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Would you visit Tynker again?</td>
<td>Yes</td>
<td>Maybe</td>
<td>Maybe</td>
</tr>
<tr>
<td>Would you play the activity you picked again?</td>
<td>Yea</td>
<td>Maybe</td>
<td>Maybe</td>
</tr>
<tr>
<td>Do you want to add anything that can help us make Tynker better?</td>
<td>Nothing, it was okay.</td>
<td>It should be written that the guidelines video is not part of Tynker.</td>
<td>Put directions voice.</td>
</tr>
</tbody>
</table>

Another observation in the first iteration was that some children were confused if the report questions were asking about Kid Crowd, the usability testing website that had the instructions for the usability test, or if they were asking about Tynker, the product under investigation. It seemed that they thought they were one website. To solve this problem, the second iteration included a page between the instructions page in Kid Crowd and Tynker in order to let the children understand that Tynker is the website that they need to test and that it is different than Kid Crowd. The page has a statement that
says: “You will go to Tynker.com to tell us how to make it better. We need your sincere opinion.” After adding this page, this problem did not occur again.

The second iteration included some other changes based on the observations in the first iteration. First, a statement was added at the beginning of the report that emphasized that the parent should read or show the choices for the child in the multiple-choice questions. Second, the important directions at the beginning of the report page were underlined, including that the parent should read the directions and type the child’s answers without changing them. Third, the last question of the report was changed from “do you want to add anything that can help us make Tynker better?” to “in your opinion, how can we make Tynker better?” This was intended to let the parent realize that the report is asking for the child’s opinion, and to directly ask the child how the developer can improve the product rather than asking him or her if he or she wants to add anything.

In session 4, although child 4 did not have a good experience with the website, her initial answers in the report were not reflective of her experience. She seemed shy to say her sincere opinion. Parent 4 explained to her that she has to “be honest” so the developers can improve their website. She said, “they want the sincere answer. They trust kids because they want to improve.” Child 4 changed her answer to one of the questions from really good to good. Parent 4 said, “good? I didn’t see you playing,” so the child said, “because I don’t know how to.” Parent 4 told her “that’s the point, they want the kid to learn, if he’s stuck and doesn’t know how to play then there’s a problem that they need to fix.” Also, when child 4 chose that the website was “somewhat fun,” the parent told her that the researcher was not the developer of the website, so the child said “oh!” Obviously, the parent realized that the child was shy to say her honest
opinion because she thought that the researcher, who was watching, was the developer of the website she was testing. After knowing that information, child 4 seemed relieved and started saying her opinion more comfortably. For example, she changed one of her answers from good to awful.

In session 5, the child 5 gave detailed answers. She was fast that her parent seemed a little behind. Parent 5 asked child 5 to repeat her answer. She did not seem like she understood what the child was talking about. After checking what the parent typed in the report and comparing it to the child’s answers, the researcher found that there were some details that were not mentioned by the parent. In general, parent 5 rephrased the child’s answers with less details. For example, one of the child’s answers were:

I didn’t really know that I was actually doing coding, like I did know, but it makes it feel like you don’t know that you’re actually learning, and it makes it really easy. Like not too easy, but like sometimes I’d get stuck but I find it at the end. And I also like that I put the steps [she means the coding blocks], and when I press play, and then if there is a mistake I would know what to add and not to add. Also, there are a lot of games to choose from, they were really fun. Just from one of the games that I’ve tried, it seems like the whole app [website] must be really fun. They have different types of games, they have games for boys, they have games for girls.

However, the parent’s answer in the report was: “It makes learning unintentional, which is something I liked. Many options to choose from. Easy to follow. I liked the colors and variety for boys and girls.”
In session 6, the parent 6’s English language was not as good as the other parents’ language. Her typing speed was also slow. Typing child 6’s answers took longer time than in the other sessions, and the child started y Dawning. She was saying her answers slowly for her parent to type them, and she seemed bored. The child helped her parent in spelling some of the words, and she explained some of the terms for her parent in Korean. In general, the parents in the second iteration read the questions for their children and typed their answers; they added the children’s opinion and encouraged them to be honest; and they showed or read the multiple choices for their children. However, some parents were slower in typing or had language difficulties. The children did not show confusion between Kid Crowd and Tynker in this iteration. The answers of the three children in the second iteration are shown in Table 5.

<table>
<thead>
<tr>
<th>Question</th>
<th>Child 4’s Answers</th>
<th>Child 5’s Answers</th>
<th>Child 6’s Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>How would you rate Tynker?</td>
<td>Awful</td>
<td>Really good</td>
<td>Really good</td>
</tr>
<tr>
<td>How much did Tynker make coding fun?</td>
<td>Just barely</td>
<td>As much as</td>
<td>Very much</td>
</tr>
<tr>
<td></td>
<td></td>
<td>possible</td>
<td></td>
</tr>
<tr>
<td>How easy was it to</td>
<td>Hard</td>
<td>Quite easy</td>
<td>Okay</td>
</tr>
</tbody>
</table>

Table 5

Children’s Answers in the Second Iteration
Table 5 continued

use Tynker?

<table>
<thead>
<tr>
<th>What do you like best about Tynker?</th>
<th>There are a lot of characters and the kids will love the animals.</th>
<th>It makes learning unintentional, which is something I liked. Many options to choose from. Easy to follow. I liked the colors and variety for boys and girls.</th>
<th>There are many games she can play. It tells you about category (beginner, advanced, ..)</th>
</tr>
</thead>
<tbody>
<tr>
<td>What do you like least about Tynker?</td>
<td>It’s confusing.</td>
<td>Nothing really. It was all fun.</td>
<td>Some games she does not understand.</td>
</tr>
<tr>
<td>How easy was it to understand the activity you picked?</td>
<td>Hard</td>
<td>Quite easy</td>
<td>Okay</td>
</tr>
<tr>
<td>Would you visit Tynker again?</td>
<td>Maybe</td>
<td>Yes</td>
<td>Maybe</td>
</tr>
<tr>
<td>Would you play the game again?</td>
<td>Maybe</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Table 5 continued

activity you picked again?

In your opinion, how can we make Tynker better?

<table>
<thead>
<tr>
<th>Clarify the instructions</th>
<th>Add more games.</th>
<th>More directions.</th>
<th>Voice to read the text.</th>
</tr>
</thead>
</table>

The six parents in the first and second iterations were asked about their opinion in letting the children type their own answers. Three parents were with the idea for different reasons, two were neutral, and one was against. When parent 4 was asked about her opinion in letting the child type her own answers, she said, “probably it’s going to be better if she did it because then she wouldn’t be faced with guilt and embarrassment because some children are just shy or sensitive they don’t want to show that they couldn’t do the test.” Also, parent 3, who actually let his child type the answers after discussing them with her, said:

I believe she should type [her answers]. First of all, it’s a practice for her. Second, it’s good to feel that she is adding value to this, this will give her confidence in herself, so she has to fill it. I believe she has to feel like independent. So, this should be part of the session.

Parent 6, who was slower in typing and was not as proficient in English as the other parents, was also with the idea. She said it is better to let the children type their own answers if they are better in typing and “more familiar with the technology” than the
parents. On the other hand, the parents in sessions 1 and 2 said that it would not make a difference whether they typed the answers or let their children do it. Parent 5 said that it would make the child stressed unless the questions are “short and straightforward”. If they were not, then it would be better to let the parent do the typing.

Based on the observations and the parents’ opinions, the researcher decided to try to let the children type their answers in the third iteration. In session 7, parent 7’s typing speed was noticeably faster than the other parents in the previous sessions. However, the child was typing slowly. When the child was done, her parent took the laptop and fixed one of her child’s grammatical mistakes. The parent said later in the interview, “when I looked at [my daughter’s] feedback I would’ve been concerned about the initial response that she wrote because it wasn’t grammatically correct, I guess.” She said that she believes it is better that the parent types the answers because the grammatical mistakes might make it difficult for the developer to understand the child’s answers.

Child 8, who only moved to the United States six months from the date of the observation, was typing using one finger and asked her parent about the spelling of one of the words. Parent 8 later said that although it was going to be faster for her to type the answers, she thinks that it is better to let the children type them by themselves. She explained:

Because I think in that way they take charge, you know, they’re the ones who have been doing it, so it’s only natural that they answer their own questions, and also it gets them used to follow those instructions and to understand what’s being asked of them, instead of me translating or trying to help her out.
Session 9 was different in that the parent 9 did not even look at the child’s answers or discussed them with him. The child’s answers were brief. In the last question, which asked the child to give his opinion on how to improve the website, he said “I don’t know.” Without looking at the question or trying to know what it was, the parent said, “just say I don’t know.” Later in the interview, parent 9 said, “I think that if I was to ask him the questions, he would’ve given me more detailed answers than him.”

When the researcher asked her why, she said:

Because he knows that his mommy wants answers, and when I ask him something, I want him to think about it. I think he would think more about what he is going to say, if he was talking directly to me.

She also said that for children with more self-confidence, like her child, it is okay to let the parent type their answers even if the children did not do well in the product they were testing. However, for the children who might be more sensitive and who might feel embarrassed to tell their opinion, it is better, in her opinion, to let them type their own answers. She also thinks that it is a good idea to give the choice to the child whether to type the answers or to let the parent type the answers for him or her.

In general, the children who typed their answers with the presence and help of their parents had more detailed answers than the children who typed their answers alone. Also, there were some grammatical mistakes in the children’s answers, but the sentences were understandable. Table 6 shows the answers of the children in the third iteration.
Table 6

*Children’s Answers in the Third Iteration*

<table>
<thead>
<tr>
<th>Question</th>
<th>Child 7’s Answers</th>
<th>Child 8’s Answers</th>
<th>Child 9’s Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>How would you rate Tynker?</td>
<td>Awesome</td>
<td>Not very good</td>
<td>Really good</td>
</tr>
<tr>
<td>How much did Tynker make coding fun?</td>
<td>Very much</td>
<td>Just barely</td>
<td>Very much</td>
</tr>
<tr>
<td>How easy was it to use Tynker?</td>
<td>A bit hard</td>
<td>A bit hard</td>
<td>Really easy</td>
</tr>
<tr>
<td>What do you like best about Tynker?</td>
<td>Playing the games</td>
<td>In the first game I liked it just because designing the actor it let me draw.</td>
<td></td>
</tr>
<tr>
<td>What do you like least about Tynker?</td>
<td>I didn’t know how to work it that well.</td>
<td>It was hard to understand what to do.</td>
<td>Nothing</td>
</tr>
<tr>
<td>How easy was it to understand the activity you picked?</td>
<td>A bit hard</td>
<td>Hard</td>
<td>Quite easy</td>
</tr>
</tbody>
</table>
Table 6 continued

<table>
<thead>
<tr>
<th>Would you visit Tynker again?</th>
<th>Yes</th>
<th>No</th>
<th>yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Would you play the activity you picked again?</td>
<td>Maybe</td>
<td>No</td>
<td>Maybe</td>
</tr>
<tr>
<td>In your opinion, how can we make Tynker better?</td>
<td>Label like if it’s a eraser and stuff</td>
<td>What I would do is to make the instructions more clear and simple. Also in the first game it said beginner but it was two hard and in the second one was intermediate it was easier.</td>
<td>I don’t know</td>
</tr>
</tbody>
</table>

In the interviews, the parents were asked if they think the children can do the whole usability test by their own, including filling the first and the second parts of the report. Some parents were skeptical about the children’s ability to complete a usability test by their own at this age. Parent 3 said that the parent’s presence is important to assist
the child who might “find some difficulties” completing the session. parent 7 was concerned about “how comfortable they actually feel with it being something new, and if no adult or parent is around, maybe they might get frustrated.” She said that the parent’s presence “could be a motivational factor” depending on the relationship between the parent and the child, and it would also help in providing a report that is more understandable with less grammatical mistakes. Similarly, parent 4, whose child had difficulties in understanding the game she was testing, said:

It depends on the website, it might be very frustrating for her that she doesn’t want to be engaged with it, like we did today, I feel like she was frustrated enough to just wanted to talk and I will do the job of typing. So, I don’t think she will be interested.

Some parents, on the other hand, believe that their children could perform the usability test by their own. Parent 8 said, “I think with the proper instructions she can,” while parent 9 assumed that since the child in crowdsourced usability testing does not do one usability test, then the more he or she repeats the process, the easier it becomes. She said:

I think he could, because there is no right or wrong answer, and he’s just testing a product, and I do feel that if he did it more, the more that he does it, the easier it’s going to become for him.

Parent 6 emphasized the importance of the parent’s presence at the beginning of each session to make sure that the product is safe for the child. In her opinion, once the safety is checked by the parent, the child can complete the test alone.
According to Hanna’s et al. (1997) guidelines, children should have a break if the session is longer than 45 minutes. In this study, the children’s participation lasted between 37 to 63 minutes, from the time they started reading the assent form until they received the gift cards. None of the children asked for a break or seemed like they needed one. Some parents said that the activity was not long enough for the children to be tired. Most of the parents said that there was no need for the break, especially at the middle of an activity. Some parents said that it would distract the child. Parent 3 said, “while the child is focusing on the game, I don’t think it’s a good idea.” Parent 7 also had a similar opinion:

I like that there was consistency, because maybe for my daughter if there was a break then she might get distracted when I come back, but I think it depends on the child and their learning capacity.

Other parents said that they do not think that a short break will affect the child’s performance or cause a distraction. Parent 9 said, “I don’t think it will affect him if he wants to go away for a minute and come back to it because he does that with homework and video games and everything else.” In general, although most of the children’s participations lasted for around 45 minutes or more, none of them was tired or asked for a break.

The last point in the findings that is related to the report part of the session is the amount of time allowed for the child to play in one activity during the session. As mentioned in Chapter 2, time pressure is one of the factors that affect the outcomes of usability testing. In this study, the children were asked to play at least ten minutes in the first iteration. However, there was no timer provided for the parents to use. All the
parents in the first iteration forgot the time their children started playing or forgot to count the time. Parent 2 said that the time should have been longer. Child 2 also wanted more time to test the product. She was aware of the timer; in the middle of the session, she asked if she still had time.

The time was extended in the second iteration; the children were told that they have 15 minutes to test the product. A timer was also added for the parents to use instead of asking them to count the time. The parents’ reactions differed in this iteration. Parent 6 said that the timer was better for them because it helped them concentrate on the activity, while parent 4 said that it limited their ability to go back and choose another activity to test when they were done with the first activity, although the parent and her child had the capacity to try another one. When her child felt frustrated with the activity that she picked, she asked if she could stop, but the parent said, “I don’t think so; there is a timer.” In the interview, she said that the timer affected the feedback she provided. She said:

I tend to write a lot usually, but I haven’t been able to really write a lot, like I was probably brief, because of the time, because I wasn’t sure if I had to pause, if I had to ask my child to wait until I finish writing, or if I just had to write like whatever, you know, what’s important.

Parent 5 was also against the timer. She said that if her child was using the website in real life, there would have not be a timer. She thinks there is no point of specifying a limited amount of time for the test. She said, “I would go with being more flexible with time,” even though fifteen minutes was not too short.
In the third iteration, the timer was removed and no limitations or instructions regarding the testing time was provided. The parents had the choice to end the session whenever they felt like they or their children wanted to end it. The time spent to test the product in sessions 7, 8, and 9 was 34, 28, and 18 minutes respectively. In session 7, the parent and the child were willing to spend even more time testing the product, but the researcher told them that they can stop if they wanted to. All the three parents in the third iteration were against setting a specific time for the session. Parent 7 said:

I like it without the time, because she got like a little anxious maybe with the first activity, it was like “oh you have” I think it was 20 seconds or something [which was part of the game itself], so she was trying to rush, so I thought because she was rushing she missed the opportunity of understanding what she could’ve learned from it. So, but I like that the rest of it didn’t have a timer because she was calmer.

She said that if there must be a timer, then it would be better that only the parent knew about it to not let the child feel anxious. Parent 9 also said that she likes the “openness” because it makes the child “feel like they are not rushed.” Parent 8 said that having a timer would not have been useful because “the whole point is that they explore it and they find things.” In general, most parents were against specifying a limited time for testing the product.

**Finishing-up-stage.** All of the children were excited to receive the gift card at the end of the session. Parents’ opinion about offering monetary rewards for children in crowdsourced usability testing websites is discussed under the parents’ concerns section of this chapter. In general, parents did not have concerns regarding the type of monetary
reward offered in this study. However, some parents emphasized the importance that this reward should be under the supervision of the parent. Parent 9 suggested, “maybe it’s better if you tell the child that it’s her money, because she made it, but it’s under the mother’s supervision.” She said, it is better when it comes from the website rather than from the parent.

Parents-Related Features

**What happened?** There are four stages that the parents went through when using the usability testing website built for this study. The first stage is the introduction. This stage includes a one-minute animated introduction video, which describes the goal of the usability testing website and how it works, as well as a sign-up procedure in which the parents created their accounts and profiles that include their information as well as their participating children’s information.

The second stage is the set-up and planning stage in which the parents selected a product to be tested, read or watched the instructions, and set up the testing room and equipment. The parents selected a product out of a list of products that are supposedly posted by instructional designers. After selecting the product, the parents read or watched the instructions that are intended to help in completing the session successfully. Then, they read the set-up instructions, which help in preparing the place in which the test will take place.

Next is the during-the-test stage in which the children joined the parents and read the assent form, completed the warm-up activity, watched the instructions, and then started trying the product. The parents’ role in this stage is to help the children understand their role by answering any questions they might have; to start recording the
session; to read the tasks for the children; to observe the children while performing the tasks; and to fill the usability testing report. The report includes two parts: The first part is supposed to be answered by the parents based on their observations, and the second part is supposed to be answered by the children, and it asked about their opinion in the usability of the product (see Appendix B for screenshots of the report pages).

The fourth and last stage is the finishing-up stage. In this stage, the parents reviewed the report before submitting it in order to add any details that they forgot to add during the session. Then, the parents were offered a ten dollars gift card as a gratitude for the time and effort they put into completing the session.

**Overview of the parents-related findings.** This section provides a summary of the findings that are related to the parents or that could help the parents provide better usability testing reports with less workload. The next section provides a detailed description of these findings.

1. Some parents could not determine if they were advanced or fluent in English language, and the number of years they lived in the country was not always a good indicator of their language skills.
2. Parents with slow internet connection could not complete the video instructions or had to watch it in parts.
3. Parents needed detailed instructions on how to set-up the testing room and equipment.
4. Having a short practice with a video example of a usability testing session seemed to help the parents in understanding their role and being prepared for the session.
5. Parents expressed their need to have the instructions in different forms: Written, and video instructions.
   
a. It was found that some parents forgot the instructions that were in a part of the video that violated one of the principles of cognitive load theory.
   
b. Parents preferred watching the instructions in shorter videos than watching one long video.
   
c. Most parents did not read the written instructions attentively; highlighting the text and screenshots seemed to help.

6. Parents expressed their need to know more about how to use the product to be tested in order to be able to concentrate on providing feedback during the session rather than trying to figure out how to use it.

7. Some of the factors that might have affected the parents’ feedback are: The parent’s knowledge about the topic of the product to be tested, typing speed, language proficiency, how prepared the parent was before starting the session, the way the report questions were written, and how much the parent encouraged the child to talk during the session.

8. Some parents found it challenging to work on two devices at the same time, but they think that this is normal since it was their first time trying such an experience.

9. Parents worked more smoothly when they were prepared for what is coming next, and when they knew who was supposed to do what and when.
10. Parents helped the children provide more sincere answers when they felt that they were not saying what they were thinking, even though some of them forgot to watch the facial expressions and the body language of the children.

11. Most parents found the observer’s role and the technical troubleshooter’s role to be the most challenging roles in the session.

**What went well and what didn’t?** This section is organized based on the stages that were described above.

*Introduction stage.* In this stage, the parents opened the homepage of Kid Crowd and watched a one-minute animated introduction video that includes the goal of the website and how it works. All of the parents scanned the home page and watched the introduction video except for one parent; parent 9 went directly to the sign-up page without reading or watching the introduction video. Two of the participating families had slow internet connections which affected the video streaming. After waiting for a couple of minutes for the video to be loaded, parent 1 decided to go to the sign-up page without completing the video.

The sign-up procedure went smoothly. The parents created their profiles by adding their information as well as their children’s information. It included the age of the parent and the child, their gender, country of citizenship, how long they have been in the US, their English language proficiency, their usability testing experience, and the child’s computer experience. One of the observations that were made in this stage is that although one of the parents has lived eleven years in the US, she was not as proficient as the other non-native speakers who have lived in the US for two years or even six months.
Also, at least two parents were not sure whether to classify their English language as advanced or as fluent, although the definition of each category was provided.

*Set-up and planning stage.* The parents in this stage prepared themselves and the testing room for the session. This stage can be divided into three parts: Selecting a product to be tested, watching or reading the instructions, and preparing the testing room and equipment.

*Selecting the product.* The parents first picked a product from a list of products that need to be tested by children of specific age and requirements. The product details included a brief description of each of the following: The product, the task, who can participate, and the reward. During the observation, parent 2 was hesitant about choosing Tynker because she did not understand the meaning of coding, but she had to choose it since it was the only product that its requirements match her child’s age. Later in the interview she suggested adding the ability for the parent to visit the websites or the products before selecting one, or letting the child select the product. She said that it is usually her older son who helps his sister, child 2, pick a game, not her. That is why she was not sure which product to choose at the beginning. She explained, “I’m not that kind of person who spends a lot of time with technology, that’s why I said she should choose [the product].” If she had another choice, she said she would have chosen another product that she understands.

While parent 2 preferred that the child picks the product to be tested, parent 1 had concerns regarding the content of the products from a cultural perspective. She suggested that the parent should be able to choose the kinds of products that can be tested by their children when he or she creates the profile. This way the parents can restrict the
children’s access to some products that might be safe for children their age, but not culturally appropriate from the parents’ standpoint.

*Watching or reading the instructions.* After reading the product’s details and selecting one that matches their children’s age and computer experience, the parents moved to the instructions page. In the first iteration, the instructions were provided only in video format. There was one video that is five minutes long. It included three parts: The parent’s role in the session, how to use Kid Crowd during the session, and some tips that were derived from Hanna’s et al. (1997) guidelines that might be helpful during the session. As mentioned earlier, the poor connection in some sessions affected the streaming of the video. Parent 1, who had problems streaming the video seemed bored while waiting for the video to load. Also, in the middle of the video, parents 1 and 2 either checked their watches or checked how many minutes were left for the video to end. Parent 2 repeated some parts of the tips’ section a couple of times. She explained, “I would prefer if there was a guidelines video and a written form of the video because I don’t like to watch, I prefer reading instead.”

During the interviews, parent 1, who had the slow internet connection, said that the instructions were clear, but she was not focusing. Parent 2 said that “the guidelines were very long, I had to read a lot and sometimes I had to go back [rewind the video] to understand.” She also said that she prefers reading, but there was only one form of instructions, which is the video. She suggested shortening the instructions.

Unlike parent 2, parent 3 said that he is “a guy who really doesn’t read instructions,” and that he tries “to figure it out very quickly.” He seemed concentrating when he watched the video, but he did not seem to read any other text in the website
attentively. He said that the instructions were not too long, and that they were comprehensive. He explained that he is a visual learner, and that is why he thinks the video was good.

Iteration 2 included the changes that were suggested in the first iteration. Written instructions were added to give the parents the choice whether to read or to watch the instructions. Also, the unnecessary parts of the instructions were removed. Again, in this iteration, the internet connection in session 4 was slow, and it affected the streaming of the instructions video. However, the parent did not click on the button that could take her to the written instructions; she said in the interview that she did not notice that there was written instructions. None of the parents in this iteration read the written form of the instructions. Although parent 5 said that she wanted to read the written instructions, she said that her child said, “no let’s go to the video,” so she watched the video instead. Parent 6 said that she noticed the written instructions, but she prefers watching over reading.

Parent 4 suggested dividing the video into three parts. She said:

I’m not sure if this is even correct, but one thing that came to my mind while I was watching the video, the introductory video, I was thinking that this is, again probably I’m wrong, but I was thinking maybe if it was divided into three videos it would be better? But I’m not sure … Because I thought like [there are] too much of instructions. I was like how can I keep up with all this.

But she said that the video was clear and she understood her role after watching it. She said, “especially when you provided the pictures, because I’m a visual person.” She found connecting what is going to happen in the report with the instructions through
visual symbols to be helpful. She explained, “I mean in the introductory video, when you use like simple objects like the bulb or like the screen, like these kinds of visual reminders are very important for people like me who are visual.”

Parent 5 also found it hard to understand all the instructions at once. She said, “I was like, will I be able to comprehend all these instructions?” But at the same time, she said that the video was “pretty straightforward and easy to follow.” She thought that dividing the video into three parts is a good idea because in the first version “you need to comprehend and process lots of information at the same time.” On the contrary, parent 6 said that the video was “not that long,” although she moved the pointer over the video when she was watching, which seemed like she was trying to see how long was left until the end. In general, the parents in the first and second iterations were confused regarding their role and what they were exactly supposed to do when the actual session started as will be discussed in the next section.

The third iteration included dividing the video into three parts. Each part was in a separate page. A worked example of a one-minute video was also added for the parents in the same page that has the video instructions that talked about the parent’s role in the session. In this example, the parents watched a part of a usability test conducted by a mother and her daughter. The video showed the screen of the daughter while she was testing a website. Parents could hear the conversation between the mother and the daughter, and they were asked to type their observations while watching the video. When the video ended, parents checked their reports with the report that was made by the mother in the video. This was intended to be like a short practice for the parents to better understand their role before starting the actual session.
During the observations, none of the parents looked at how much was left in the videos. Two of the videos were one minute and twenty-eight seconds long, and the third video was two minutes and eight seconds long. More than one parent in the three iterations nodded their head when the meaning of the visual icons that are used later in the report were explained in the video. Parent 7 rewound a part of one of the videos. When asked if the video was too long or if it included too much information, she said:

One of the videos, it felt like a lot of information, so I just went back and listened, maybe rewind 40 seconds, and then listened again to make sure I understood, but, so that maybe a little bit, but it wasn’t like something that would detour me from watching.

Also, none of the three parents in this iteration read the written instructions although they all said that they noticed the button, and one of them said that she prefers reading the instructions. Parent 8, who said that she prefers reading, said that she was “curious about the video” that is why she watched instead. Parent 7 said:

I prefer the video because it’s straightforward, but I have used for work like applications, where there’s like “oh, read the transcript or watch the video,” and sometimes I’d read the transcript if the video is too much, but I felt like “so, I will watch the video first and if the video had too much information, I will read it,” but I like your videos so I didn’t feel like I needed to read.

Parent 8 suggested that it was going to be better if each part of the instructions was followed by the corresponding part of the session that the parent was supposed to work on rather than watching all the videos and then applying what was in them. She said that it is better this way because otherwise “you lose the flow or the fluidity.” She
also said that it was good that the tips video was at the end right before starting the session so that the parents do not forget what was mentioned in it.

Regarding the short practice, parent 7 said, “I found the short practice extremely helpful in learning beforehand what I needed to do and allowing me to practice before starting the session with a test run.” Parent 8 said that it would not have made a difference if it was not there, but then she remembered that she had to compare her answers with the answers of the parent in the video, and she said that she liked that. Parent 9 also said, “I think that was helpful. It let me know what I was supposed to do.” When the actual session started, the parents knew what they were supposed to do, unlike the parents in the first and the second iterations.

The last finding in this section was that parent 2 and parent 9 mentioned that they only remembered the part of the instructions that said when your child asks you how to do a task, redirect his or her question by asking “how do you think you can do it?” But they did not remember the rest of the sentence that said do not offer any suggestions or hints until your child is stuck. They remembered the second sentence only later during the interview. Parent 2 said:

I remember now it did say that I have to help her when she asks questions, but when she was doing it, I thought I shouldn’t interfere with her when she plays in the program, but then I remembered actually that I should help her.

Parent 9 also said that one of the problems that she faced during the session was that she wanted to guide her child more than she was allowed to. When the researcher reminded her that the instructions said that the parent can help the child only when he or she is
stuck, she said, “you know, I remember them saying it now, but I didn’t remember it when he was doing it.” She explained the reason why she forgot that specific part:

I guess what I remembered was where you guys said if they ask you a question, you basically give them a question back. So, I think that’s what was stuck in my head the most, and I think the very last part of that, those instructions where you said, “if they get stuck,” I think you ended the sentence in the word stuck, and I think all I remembered was what I saw, because I think you had pictures to go with that, you said “if they say this” you had a picture, and that’s what I had in my head. So maybe either put a picture there for “if they get stuck,” or put “you can help.”

What parent 9 meant was that for each piece of instructions in the video, a new animation appeared, but for that specific instruction that said, “do not offer any suggestions or hints until your child is stuck,” the animation of the previous sentence was still there on the screen because the researcher considered them as one piece of instructions. Based on the parent’s explanation, this is why some parents did not remember the second part of the sentence. Figure 3 below shows how the animations were changed for each piece of instructions, but not for the part that the parents forgot.
Encourage your child to verbalize his or her thoughts. For example, if your child seems like thinking, try to get a clarification of what he or she is thinking.

Always provide positive feedback even if your child could not complete a task.

When your child asks you how to do a task, redirect his or her question by asking “how do you think you can do it?” Don’t offer any suggestions or hints until your child is stuck.

Figure 3. A problem with the timing of the visual and the auditory information.
Preparing the testing room and equipment. In this part of the session, the parents read a list of instructions that were derived from Hanna’s et al. (1997) guidelines. These instructions were related to preparing the room in which the session will take place. After reading the instructions, the parents filled a checklist to make sure they followed all the instructions. In case they could not follow all the instructions, a text box was provided for them to explain the situation. Parent 7 said that she liked the checkbox. “It’s like okay, you watched this video, but did you do these things? So, I really like the process,” she said. One of the observations that were made was that child 2 was sitting behind a light source, which affected the quality of the recording, and made it hard to see her face expressions. The second iteration included this point as one of the set-up instructions.

During-the-test stage. The parent’s role in this stage was to answer the child’s questions before starting the session; to start recording the session; to read the tasks for the child; to observe the child while performing the tasks; and to fill the usability testing report. This stage can be divided into two parts: Recording and starting the session, and observing the child and filling the report.

Recording and starting the session. After preparing the testing room, the children read the assent form, completed the touchpad tutorial warm-up activity, and watched the instructions. Next, the parents were supposed to follow the instructions to start recording the session. Most parents had problems following the instructions in the first iteration although there were screenshots of what they were supposed to do. They were especially confused about how to record the session, and what to do next.
In the second iteration, red arrows were added to the screenshots to make it easier for the parents to spot where they were supposed to click or find something. Also, the parents were noticed to scan the text instructions that accompanied the pictures. Therefore, the statements that were most important were changed to red with bold font. The page that had the recording instructions was divided into three steps in order to help the parent know when to do what because in the first iteration, children went to Tynker and started using the website before knowing what was the task. Therefore, after the children completed the instructions video, a stop sign appeared on their screens with a note that says please remind your parent to start recording the session, and wait for him or her to read the first task for you, then go to Tynker. This was intended to allow the child to remind the parent about the order of the steps, and in case the parent is holding the child’s laptop, he or she would not miss the instructions that are on his or her laptop. In other words, the instructions that were on the parent’s laptop were briefly duplicated on the child’s laptop. Parents 4 and 5 followed the recording instructions easily, while parent 6, who had some language difficulties, did not. Child 6 read the note that says that the parent had to record the session and started helping her parent in the recording steps.

In the interview, four out of nine parents said that the most challenging role in the whole experience was the technical troubleshooter because they are not good at technology. In general, it seemed that the fact that the parent had to read the instructions on one laptop and apply them on another was problematic. Parent 5 and parent 8 said that it was hard to read the instructions in their device and apply them in the child’s device. Parent 8 said, “you do go step by step, but then it’s kind of weird going from one computer to the other, not that it’s hard, but it’s just you kind of lose flow.” However,
parent 5 thinks that being confused due to being “divided between two laptops” is normal at the beginning. She said that normally “we take time at the beginning and then everything goes smoothly, regardless of the number of the instructions or tasks.” When parent 9 was asked if working on two laptops was hard or distracting, she said “no, but I think that’s because I’m used to do stuff like that” since her job is an IT specialist.

As mentioned earlier, the parents in the first iteration were confused regarding what to do next. All the three children in that iteration started using the website before listening to the first task that the parents were supposed to read. The parents did not know that they had to read the tasks and type their observations while the child is using the website. Also, parents 1 and 2 did not know that the report was divided into two parts; the first part was for them to read the tasks and type their observations, and the second part was for the children to answer. Both participants suggested that the parents should be informed ahead of time that there are two parts in the session, and that they should know their roles in each part for them “to be ready.” Parent 3 also suggested to add a note right before the report’s page to remind the parents that they should read the tasks for the child and type their observations.

A note in the page that is before the report’s page was added to let the parent know that there are two parts of the report, and what exactly is going to happen in each part. The titles of the report pages were also changed to Report (1/2) and Report (2/2), and the buttons that lead to each part were labeled as follows: Next (Report 1/2) and Next (Report 2/2). A stop sign was used on the children’s interface, as explained above, with a note that asks the child to remind his or her parent to record the session and read the tasks. This clarity of what is going to happen next, and who should do what and when
seemed to help the parents managing and starting the session properly in the second iteration.

*Observing the child and filling the report.* The report in the first iteration had five tasks; three were followed by multiple-choice questions that help the parent tell the instructional designer exactly how his or her child responded to each task, and two tasks were followed by open-ended questions to allow the parent to either type the child’s answer or to type any usability problems the child has faced during his or her use of the product. Also after each of the three tasks that were followed by multiple-choice questions, there was a text box that is labeled with the word “notes.” The text boxes were intended to be as a space for the parent to type any usability problems that were faced while the child is performing the three specific tasks. For example, one of these tasks asked the child to pick an activity that matches his or her coding skills. The task was followed by the following question: Did your child notice the beginner, intermediate, and advanced categories? The parent could choose one of the following answers: Yes, no, or yes, but he or she did not pick an activity that matches his or her coding level. A text box was provided under the choices, and it was labeled with the word “notes” so that the parent can type his observations in case anything happened that is not in the choices.

It was noticed in the first iteration that some entries in the text boxes that were under the multiple-choice questions were brief or not helpful. Even the entries in the text box that was under the last open-ended question, which asked the parents to watch the child playing the activity for at least ten minutes and type their observations, were also brief. Parent 1 said that her answers were brief because her child had a good experience and did not face any problems. She said, “I wasn’t so informative because it was based
on her experience not mine.” Parent 2 said, “I thought that I shouldn’t write everything in detail, I don’t know because it said “notes,” so I thought I will just write notes, brief notes.” Parent 3 said that “because it is the first time” he participates in a usability test, this might have affected the quality of the feedback provided, especially that he did not read the instructions, which caused him to not start the session by reading the first task and then typing notes and so on. Being not prepared for what he was supposed to do affected his report in his opinion. His answer to the first question shows that he was typing all of his observations in the first text box rather than moving to the next task. In general, all three participants did not start the session properly as they were supposed to do. As mentioned earlier, they all suggested that the parent should know what exactly is coming next in order to be prepared. Table 7 shows the parent’s reports in the first iteration.

Table 7

Parents’ Reports in the First Iteration

<table>
<thead>
<tr>
<th>Questions</th>
<th>Parent 1’s Answers</th>
<th>Parent 2’s Answers</th>
<th>Parent 3’s Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type the child’s answer: What</td>
<td>I think it is pretty</td>
<td>It is about teaching</td>
<td>Coding game 3 stars</td>
</tr>
<tr>
<td>do you think the website is for?</td>
<td>cool, has a lot of games, are their grades? Oh, there are grades two and</td>
<td>children how to use the computers. I am not sure. I think it’s about teaching children, some</td>
<td>for coding. I need to add one more block.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

She does not like brick breaker the level button is not clear.
Table 7 continued

<table>
<thead>
<tr>
<th>How long did it</th>
<th>Ten seconds or less.</th>
<th>Ten seconds or less.</th>
<th>Ten seconds or less.</th>
</tr>
</thead>
<tbody>
<tr>
<td>take your child</td>
<td>to find the</td>
<td></td>
<td></td>
</tr>
<tr>
<td>grades button?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes. I had to reread it. I gave her a hint so that she can find the ages. she did not realize that 3-5 means all children at these ages can play the same games until I
explained for her. She thought a specific game would be assigned to 4 graders.

| Did your child notice the coding levels? | Yes. | No. | No. |
| Notes. | Yes, she picked the beginner level. | She chose one and then she went back and she chose another simpler skin I guess. |

| Your child seems like s/he is: | Engaged. | Engaged. | Having fun. |
| Explain. | She was very interested, having fun and enjoying the activity. | She says it is not very difficult to use and play. |

| Did your child face any technical problems while | No. | I don’t think so. | Not really. |
In the second iteration, the word “notes” was replaced with a statement that asks the parents to type their observations in detail. It also included some examples of the usability problems that the parent was supposed to record once noticed. Filling the text boxes was unrequired, meaning that the parents could go to the next page without typing anything in the text boxes. Also, it was mentioned above that in this iteration the parents were informed exactly what is coming next; the titles of the report pages and the buttons that lead to them were changed to clarify how many pages there were in the report, and a note was added before the report page to inform the parents about what they were supposed to do in each report page. However, although the parents in this iteration entered more details, and the procedure went more smoothly, the text boxes that were under the multiple-choice questions seemed to confuse the parents; instead of moving to the next task, the parents would type their observations in the text box and watch the child continue using the product. During the observation, when parent 4 did not read the next task, the researcher told her that she was supposed to read it. She said that she did not know she had to move to the next task because of the text box. Later in the interview, she was asked whether the statement inside the text box that gave her some examples of usability problems was helpful, she said that it was helpful, but she still did not type in detail because of the timer. She said:
I was probably brief, because of the time because I wasn’t sure if I had to pause, if I had to ask my child to wait until I finish writing, or if I just had to write like whatever, you know, what’s important.

Another observation that was noticed in this iteration was that parent 6 did not fill the report while her child was using the website. She waited for her until she was done, and then she used her child’s help to fill the report. The child told the parent what to type and also helped her in spelling some of the words. She also explained some things to her parent in Korean. Later in the interview, the parent was asked if the language was an obstacle during the session. She said, “I don’t think so. If my English is better than now, the result is the same. The website was not required difficult English as completing the session.” Table 8 shows the parents reports in the second iteration.

Table 8

*Parents’ Reports in the Second Iteration*

<table>
<thead>
<tr>
<th>Questions</th>
<th>Parent 4’s Answers</th>
<th>Parent 5’s Answers</th>
<th>Parent 6’s Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type the child’s</td>
<td>I think it’s coding.</td>
<td>There are various</td>
<td>I think it’s about</td>
</tr>
<tr>
<td>answer: What do you think the website is for?</td>
<td>But I also think it’s about math and the solar system. So, I think this website is for school. What I</td>
<td>options, I could do this the whole day. I liked the games.</td>
<td>math</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Filled by her child)
Table 8 continued

<table>
<thead>
<tr>
<th>Question</th>
<th>Response 1</th>
<th>Response 2</th>
<th>Response 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>How long did it take your child to find the grades button?</td>
<td>Ten seconds or less.</td>
<td>Ten seconds or less.</td>
<td>Ten seconds or less.</td>
</tr>
<tr>
<td>Did your child notice the coding levels?</td>
<td>Yes.</td>
<td>Yes.</td>
<td>Yes.</td>
</tr>
<tr>
<td>Your child seems like s/he is:</td>
<td>Frustrated.</td>
<td>Having fun.</td>
<td>Having fun.</td>
</tr>
<tr>
<td>Did your child face any technical problems while using Tynker or while playing in the activity?</td>
<td>Maybe the instruction for the game was not clear enough for her. I think the colors matches the proposed theme. It was easy to understand the directions as much. She also was bored and unsure while reading the instructions. She didn’t follow and she seems to learn through trial and error. Almost by the time</td>
<td>Maybe the instruction for the game was not clear enough for her. I think the colors matches the proposed theme. It was easy to understand the directions as much. She also was bored and unsure while reading the instructions. She didn’t follow and she seems to learn through trial and error. Almost by the time</td>
<td></td>
</tr>
</tbody>
</table>
Table 8 continued

she is done reading
the instruction, she
decided she is a
beginner.
She was frustrated by
the game she chose
and quit the game
after many tries.

In the third iteration, the text boxes below the multiple-choice questions were removed. Only two text boxes were in the report, and they followed the open-ended questions. In the first one, the parent was supposed to type the child’s words when first opening the website. In the second one, which is at the end of the report, the parent was supposed to watch the child while he or she is playing in the activity and type any observations that are related to the usability of the product. The parents’ reports in the third iteration were noticed to be more detailed. The sessions also went more smoothly, and less confusion was noticed in terms of starting the session with the first task and moving from one task to another. It was also noticed that the parents in session 7 and 9 in this iteration had better typing skills than all of the other parents in this study. This allowed them to watch their children while typing on their devices without having to keep looking at their own screens. It is also important to mention that none of the parents in this study had experience with coding except for parent 9. However, although she had
experience with the type of games her child was playing, she did not interfere or try to guide him when he was stuck. On the contrary, parents 7 and 8 were having conversations with their children while playing in the activities they picked, although they did not have any coding experience. Table 9 shows the parents’ reports in the third iteration.

Table 9

*Parents' Reports in the Third Iteration*

<table>
<thead>
<tr>
<th>Questions</th>
<th>Parent 7’s Answers</th>
<th>Parent 8’s Answers</th>
<th>Parent 9’s Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type the child’s answer: What do you think the website is for?</td>
<td>Site for kids to play video game and all games. The bar is different levels of games. The side is to play games like Minecraft and build a robot with all types of games. Test questions at the beginning before she plays the game. This is a game where you earn minutes of playing games.</td>
<td>Animated games. Different games. Mostly with animals or aliens. They are games for kids. Hour of code is like &quot;time to play.&quot;</td>
<td>He thinks it's a video game website and didn't notice at all that it was a site for coding. He recognized a few of the titles and got excited about playing them. Never figured out that it was a coding website from just looking at it.</td>
</tr>
<tr>
<td>How long did it take your child to find the grades button?</td>
<td>More than 60 seconds, or she didn't find it.</td>
<td>More than 60 seconds, or she didn't find it.</td>
<td>Ten seconds or less.</td>
</tr>
<tr>
<td>Did your child notice the coding levels?</td>
<td>Yes.</td>
<td>Yes.</td>
<td>Yes.</td>
</tr>
<tr>
<td>Your child seems like s/he is:</td>
<td>Engaged.</td>
<td>Frustrated.</td>
<td>Engaged.</td>
</tr>
<tr>
<td>Did your child face any technical problems while using Tynker or while playing in the activity?</td>
<td>[Daughter's name] was engaged. She did not notice the amount of time she had. She only had a few seconds and was confused about how to start making the monster. Overall, she liked the activity but was confused how to start. [Daughter's name] continued to play the game and won.</td>
<td>A little distressed at first and very frustrated later. She did not seem to find easily what to do with the things shown. It took her some time to find the codes. After a while, she found the codes but she was not able to make her character move. She was looking through the instructions on the left side of the screen so he just started clicking things and moving them around.</td>
<td>He did, he never noticed the character button and was able to go in and change his character. After he was done</td>
</tr>
</tbody>
</table>
points. She was excited at the feedback the game gave her as she continued to play. She enjoyed reading the instructions out loud. When she clicked on next, the game took her back to the same activity. [Daughter's name] was eager to go to the next level. She likes the monster with the tablet rather than seeing the circle showing the game loading. She thought it was cool. [Daughter's name] like the game with the thimble. It was challenging and required [Daughter's name] to think more looking for help and could not get that to help her solve her problem. She knows she has to make the character do things really know what he was supposed to do does not know how to get the code to work. She got frustrated.

characters, he did not notice the save button and just hit the back button to exit out of
Table 9 continued

about the number of steps and actions required to get the candy. She became tired after a few attempts and wanted to try a new game. She likes the site so far. The next game she chose was Spin Draw and she really liked the idea of creating the background of the game. She is looking at different background and she selected the background of treasure island. [Daughter's name] would hit the buttons too fast or too many times and the background. The games appeared as if
Table 9 continued

she had to start over
but then the new actors
came up. She liked the
characters. She stated
they look like
characters from movie
star planet. She enjoyed
building the character
that resembles her
features. I believe she
likes the diversity of
the game with skin
color, hair, facial
features and clothing
options. She looked for
options that she felt
looked like her. The
screen seems difficult
to scroll down as she
was building her
character options.
Some of the items for
the did not have a
description of what it
was used for. She was
confused about the
spinner option.

Most of the parents did not use the help or the support provided in the report
tables, except for the parents in sessions 1 and 4. Parent 4 said that the support provided
when the parent hovers over the lightbulb “was really helpful.” She also said, “in the
introductory video, when you use like simple objects like the bulb, or like, you know, the
screen, like these like kind of visual reminders are very important for people like me who
are visual.” Parent 1 said that “it was very thoughtful,” and that she “thought of it as one
of the kids’ rights” because some of the support hints remind the parents how to deal with
the different situations that might affect the child. Most of the other parents said that they
did not need to use the support or the help page.

In the interviews, when parents were asked about the most challenging role they
took in the session, six of them said that it was the role of the observer. Parent 1 said that
observing the child was difficult because she had to observe her body language and the
facial expressions as well. Parents 2, 4, 7, and 8 said that it was challenging because of
their lack of knowledge about coding and coding games, which is the topic of the website
that the children tested. Parent 2 said, “I was like, what is she doing really? Is she doing
it right or wrong?” Parent 4 explained:

Probably the most challenging role I faced is when I felt she wanted me to explain
the game for her, so I felt like oh, okay, because I thought she will understand the
game by reading the instructions, that’s why I kind of told her familiarize yourself with the game before you play, before I hit the play button [for the timer], but then when she couldn’t understand, I felt frustrated because I felt oh, okay, I had to pay attention.

Also, parent 7 had a similar answer. She said, “I don’t know a lot about coding, and I don’t know that I’m the most technically savvy when it comes to apps and things, so I’m probably at the beginner level too.” Parent 8 said that if she did not know from her readings that coding is for programming, she “would not understand anything at all.”

Parents 7 and 8 suggested that the parents should know more about the product to be tested before the session starts so they can better support their children and to be able to provide a better report. Parent 7 suggested:

[If there were] some learning objectives that the parents can read, for the different games … maybe if this is in the parent’s part, then I would know kind of overall, “these are the games at this level that your child should take away, like the beginner level will teach them these things, or the intermediate, and advanced levels will teach them different things,” and then maybe add some challenges maybe that I may encounter on those levels, and maybe what parents can do to help, support, or I don’t know.

Parent 7 had a similar suggestion:

It would be best if the parent gets a little more information about what is the game about, the kids going to test, and how you play it, to sort of like understand it better, so maybe you know a little tutorial video, a minute or so, where you can actually see literally how you do it, but why you’re doing it. Maybe a short video
of what do you have available in that game, and how to use the tools, well, I think this is the whole purpose of testing, right? But maybe for the parent to know beforehand then we can guide the kids a little better.

Parents 5 and 8 felt that it was challenging because they were divided between two laptops, and they had to do multiple things at the same time. Parent 5 said, ‘you need to manage between different things, observe and write feedback and follow instructions, and see what’s next. The multi-task thing.” Also, parent 8 said that what was challenging is “trying to understand the game, at the same time being an observer, at the same time recording the comments.” She added, “because I’m new at all of those at the same time, so I have to be focused at remembering different tasks.” Parents 5 and 8 thought that the process will be easier with time. Parent 8 said, “I think it will get easier if I repeat it, and also if I know the game beforehand.”

On the contrary, parent 9, who is an IT specialist and has previous experience with coding and with working with more than one device at the same time, said that the observer role was challenging because she “wanted to help him more” than she was allowed to. She said, “I mean it wasn’t difficult, but it was frustrating a little bit just because I wanted it to be like “hey, do this, or maybe try this.”” As explained earlier, she thought that she was not allowed to interfere and help her child when he is stuck, and that made the observer’s role more frustrating for her.

The last point in this section is the parent’s ability to include the facial expressions and the body language of their children in their observations. Most parents said that they did take into consideration the facial expressions and body language of their children when they were filling the report even though it was not always intentional.
That means that even though they might have forgotten the instructions regarding observing the child’s facial expressions and body language, they did include them in the report. Parent 5 said, “I didn’t really pay attention to her face, but as a mother, you know what she’s saying, is it really reflective or she’s just saying these. I really didn’t feel any discrepancies between what she said and [what she felt].” Other parents had similar responses. Parent 8 said that she “completely” forgot to watch her facial expressions, but she said that she “could feel her stressed at some point,” and that she stated in the report that “she was kind of distressed at the beginning and then she was frustrated.” When parent 4 was asked if she took into consideration her child’s face expressions, she said, “I noticed that. That’s why I tried to follow her answers with other questions because I felt like her answers were different than what I’m feeling.”

**Finishing-up-stage.** In this stage, the parents were supposed to review the report before submitting it, and then they receive a reward for their participation. The review step was intended to help the parent type any details they missed during the session if they did not have enough time to type everything they noticed while they were observing their children. However, it seemed from the observation that none of the parents have reviewed their answers. Most of them clicked on the button that takes them to review their answers, but they did not actually review their answers. Parent 2 explained, “I don’t think there is a need to review my answers because usually after I write a sentence I reread it.”

**Parents Concerns**

The usability testing tool that was developed in this study is a crowdsourcing website. This type of remote usability testing was picked because it is a promising new
approach to usability testing that might be able to solve the problems of traditional usability testing approaches, but, as far as we know, it was only tested with adult users. Crowdsourced usability testing websites are a particular type of remote usability testing websites. The main difference between the two is that the participants in a regular usability testing website would normally be invited to participate only once, while in the crowdsourced website the participants would have an account where they can log in and check the available products that they can test, and they can earn some kind of reward or benefit for each product they test. Therefore, it was worth investigating the parents’ concerns regarding their children’s participation in remote usability testing in general and in crowdsourced usability testing in specific, which might inform the guidelines of designing such platforms.

**Overview of the findings related to the parents’ concerns.** This section summarizes the findings that are related to the parents’ concerns regarding children’s participation in crowdsourced usability testing. The next section provides a detailed description of these findings.

1. Parents’ concerns regarding safety and privacy:
   a. The parents’ approval for their children to participate in a crowdsourced usability testing website without the help of the parent depends on how trusted and safe the website is.
   b. The parents’ concerns regarding safety and privacy depend on whether they are with the children while using the website or not.
   c. Parents are mostly concerned about two issues: The privacy of identity, which is mainly represented by the parents’ concerns regarding
videotaping their children, and the appropriateness of content, which is
represented by the parents’ concerns regarding the appropriateness of the
products to be tested in terms of age and culture.

2. Parents’ concerns regarding how long and how often their children can participate
   in usability testing:
   a. Most parents perceived their children’s participation in usability testing as
      a positive experience.
   b. Some parents believe that a usability testing session should not be longer
      than an hour for a comfortable experience for the child and the parent,
      while the others believe it should not be longer than fifteen minutes.
   c. Most parents are against setting a timer for the usability test. They believe
      it has negative effects on the child’s and the parent’s performance.
   d. Most parents are willing to allow their children to frequently participate in
      crowdsourced usability testing, while some emphasize the importance of
      controlling the children’s participation so that it does not become
      addictive.

3. Parents’ concerns regarding the reward that the children obtain from participating
   in crowdsourced usability testing:
   a. Parents do not have concerns regarding the regular monetary rewards as
      long as it is under the parent’s supervision.
   b. While some parents believe that the mini-payments that are paid for mini-
      tasks can engage the children into doing more tasks, other parents are
afraid that this type of reward could be addictive, or that it could make the children care about the money more than the learning part of the process.

c. Other types of rewards have been suggested as alternatives.

4. Other concerns:
   a. Some parents are concerned about the children’s feelings when they are constantly being provided with negative feedback from testing defective products.

   **Safety and privacy.** Two main safety and privacy concerns were identified regarding children’s participation in remote or crowdsourced usability testing. These two concerns are the privacy of identity and the appropriateness of content.

   **Privacy of identity.** Most parents had concerns about videotaping their children during the usability testing session. Parent 7 said, “I wouldn’t want her to be recorded on someone’s site ... I would be fine with [the other types of recordings] just not recording like her, her face I mean.” Parents 2, 4, and 9 did not see a need for videotaping their children, especially that the parents are supposed to be present during the test, and in their opinion, their presence compensates for the need of this kind of recording. Parent 4 said:

   I would understand the need for videotaping the child if he is doing the experiment by himself because the instructor will want to analyze, you know, the facial expressions and all of that, but if the parent is next to him and trying also to give like a second opinion or trying to support, then why would [it be needed]? …

   From my own kind of judgement, I would be interested in taking my kid to do this experiment, but I wouldn’t be comfortable if there would be videotaping.

   Parent 9 had a similar opinion:
If they asked them to record their face, probably I would be like “that’s kind of weird,” but if they just asked to record like their voice or, you know, their screen, whether there was clicking and stuff, I’d probably be fine with that, yeah…

[It’s unnecessary] because I couldn’t hardly tell that much from his face. Yeah, and they’re asking us to already look at them.

Parent 2, who shared the same concerns as parents 4 and 9 in that her notes were enough for the instructional designer to obtain the required feedback, also expressed her concerns about her child’s preference:

I think my daughter wasn’t [comfortable with videotaping] either because she doesn’t like anyone to videotape her or take photos of her … Recording the screen is a good idea. I don’t have a problem with the audio, but I don’t know about [my daughter] actually. Because I know that she doesn’t even let me videotape her.

Besides the belief that videotaping the children during the session does not have any benefits for the instructional designer, parent 8 explained the reason why it is not acceptable to videotape the children during a usability testing session:

Just about what will happen with those recordings, you know, and in general I try not to have my children being exposed to any recordings of any time, just as a precaution, wouldn’t put on Facebook [for example], but just maybe because also I’m a journalist and I’m a little paranoid of those things.

When asked about screen recording and audio recording, she said “it’s a lot better,” but the concern was about recording the face, or “the identity.” Six parents expressed their concerns regarding videotaping their children’s faces, but they all were fine with the
other types of recordings, including audio recording the child’s and the parent’s voices and the screen recording.

On the other hand, parents 3, 5, and 6 were fine with all kinds of recordings, including videotaping the child’s face, as long as the website they are dealing with is trusted. When asked about their concerns about the three types of recordings, parent 5 said, “yeah, I don’t have any problem,” while parent 6 explained the importance of recording the face of the child for the instructional designers by saying that “they have to know that she really used this website.” When she was asked about her concerns in general, she said, “if the information is not going to spread out, it’s fine.” Parent 3 had similar opinion. He explained:

Yeah, it’s acceptable, and as I told you, as long as I know who I’m dealing with, I believe it’s acceptable. It’s the same thing when you record yourself and you have a camera in your mobile and you record your children in your mobile, so the same thing because you trust that it’s not going to be hacked or something like that. So, the same idea.

The safety and privacy concerns also occurred during the observation. Parents 1, 2, and 4 expressed their desire in not being video recorded. Parent 2 constantly reminded her child not to move the laptop’s screen towards her so that she does not show in the recording. Parents 1 and 5 started speaking in a very low voice once they started recording the session, although parent 5 was surprised when she was asked later in the interview why she lowered her voice. She said, “interesting, I didn’t notice … maybe because at the beginning you’re not used to it, but then it’s normal.” She also added, “I
guess it’s the shyness, at the beginning” and that it was “unconscious,” which indicates that it might or might not be related to the safety and privacy issues.

In general, most parents had concerns regarding the privacy of information and identity represented in videotaping the session. The reasons indicated by the parents were that they did not see a benefit from videotaping the child during the session because their presence was enough to provide the same information that the instructional designers would obtain from the video recordings. On the other hand, some parents acknowledged the need for the instructional designers to make sure that a child has actually tested their product. However, they emphasized the need for the parent to make sure that the information would not spread, and that the usability testing website is trusted or certified by an authority, as suggested by parent 3, in order to accept videotaping their children. Parent 7 also suggested providing a disclaimer when she said, “wondering maybe if the recording, if this going to give someone access to like my computer screen … So maybe if there’s some kind of disclaimer or something in there.”

**Appropriateness of content.** The second main concern regarding the safety and privacy of the children in crowdsourced or remote usability testing is the appropriateness of the content provided in the products to be tested. This concern can be classified into two categories: Age-appropriate content and culturally-appropriate content.

When he was asked if his child can participate in crowdsourced usability testing by her own, parent 3 said:

I would let her do that, but I have to trust the website … I cannot trust something that comes from YouTube, for example. I don’t know who is the issuer, what kind of games or what kind of things they are putting. The same thing when you check
for your child when they use the internet, I believe it’s the same thing. If she’s
going to go to a website that I know, I wouldn’t check for her, I will let her go,
but if she’s going to a website that she may find good things and bad things, here I
have to watch her, for example like YouTube or Google, where she can find bad
and good things.

A similar answer was provided by parent 5. She said, “in general, I need to make sure
that the content is appropriate.” When she was asked if it will help if the parent could
check the product before letting the child test it, she said, “since it’s educational, I don’t
think I will need it, but I would go with this in YouTube for example or in any movie.”
She believes it depends on the type of products, if the products to be tested are all
educational products, she “would go %100 trust.”

Parent 6 said that she would let her child perform the usability test by herself, but
she has to be with her at the beginning of each session, and that is because she needs to
“check the website if it’s safe or not.” Parent 1 also mentioned that the website has to be
free of advertisements. When the researcher said that the list of products that are
provided in the usability testing website go through online safety and privacy check for
children before being accepted in the list, parent 1 explained that it is not only about age-
appropriateness. She said, “they might be suitable for kids, but not culturally
appropriate.” She suggested, “if I can do something with my account to choose, like one
of the options when I do my profile, to choose what my kids get, so when they sign in
again if they are using it under my name, so I can know what they are going to test.”

**Time and frequency.** Since usability testing is originally conducted in a lab with
the presence of a test administrator, it is not the parent’s responsibility to determine the
length of the session. However, in remote usability testing, it is important for the
designer to specify the appropriate length for each session taking into consideration the
parent’s and the child’s capacities. In addition, in regular usability testing the participant
is supposed to test one product. However, in crowdsourced usability testing, the
participant has an account in which he or she can visit the website frequently to conduct
several usability testing sessions. Therefore, the length of the usability testing session
and the frequency of testing the products are investigated from the parent’s point of view.

Length of the usability testing session. From the participants’ opinion and
experience as parents, most of them believe that the maximum length of a remote
usability testing session with children should not exceed 40 to 60 minutes for a
comfortable experience for both the parent and the child. Parent 3 explained, “she’s
learning from it. So, I believe 45 minutes to one hour is good.” He further explained why
in his opinion it should not exceed an hour, “I think it’s good to have one hour maximum
for her, so she doesn’t feel bored about it or that it’s something she has to do, but one
hour is fair I believe.” Parent 4, whose child spent 34 minutes testing the product,
believes that it depends on the child. She said:

I feel that my kid was willing to spend more time to understand it, but this also
depends on the child and his attention span, but for [my daughter], I felt like she
wanted to figure it out, so she wanted to understand it. So, it was okay for her to
go longer.

When asked about her preference as well since she has to be present in the session, she
said, “I feel that as long as the kid is engaged, then I wouldn’t mind going longer … I feel
that if there is no limitation, I would go back and choose another activity.” Parent 7 also
thinks that it depends on the child, “I think it shouldn’t exceed maybe an hour, depending on the age, maybe an hour and a half.”

On the other hand, parents 1, 2 and 8 were more conservative with a maximum length of 20 to 30 minutes, 15 to 20 minutes, and 15 minutes or less, respectively. Parent 8 explained:

I think it shouldn’t be more than fifteen minutes. I think that if the kid really likes, I mean, once we already know the whole procedure, right? So, we already know everything, we’ll be like “okay, let’s try this one” … at most I would say fifteen, because really it could be even less than that, five or seven minutes.

She then explained that there should be control over the child’s use of technology to leave a space for the other things that can be experienced.

Although parent 2 said that the length of the session should be between 15 to 20 minutes, her daughter, child 2, who seemed to be enjoying playing in the product that she was testing, was upset when the parent asked her to stop using the product. She actually asked her parent if she could pick another activity. Children 6 and 7 spent more time by testing two activities rather than one, while child 3 said that she does not want to test another activity when she completed the first one. Child 4, who was trying an activity that has several usability problems that prevented her from actually playing the game, asked her parent if she could stop after around 30 minutes of playing. Her parent said later that her child was willing to continue the session and that she was willing to test another game if the time was not limited.

Table 10 shows the time spent testing the product starting with the time the child goes to Tynker until the parent starts asking the first question in the second report. The
minutes provided does not include the time spent answering the report’s questions. In the first iteration, the children were asked to play at least 10 minutes. In the second iteration, the children were told that they have 15 minutes to test the product, and a timer was added for the parent. In the third iteration, the timer was removed and no limitations or instructions regarding the testing time was provided.

Table 10

*Time Spent Testing the Product*

<table>
<thead>
<tr>
<th>Iteration Number</th>
<th>Session Number</th>
<th>Time Spent Testing the Product</th>
<th>When Did the Child Stop?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iteration 1</td>
<td>Session 1</td>
<td>14 minutes</td>
<td>The mother asked the child to stop.</td>
</tr>
<tr>
<td></td>
<td>Session 2</td>
<td>18 minutes</td>
<td>The parent asked the child if she was done.</td>
</tr>
<tr>
<td></td>
<td>Session 3</td>
<td>13 minutes</td>
<td>The activity was completed.</td>
</tr>
<tr>
<td>Iteration 2</td>
<td>Session 4</td>
<td>34 minutes</td>
<td>The child asked to stop.</td>
</tr>
<tr>
<td></td>
<td>Session 5</td>
<td>18 minutes</td>
<td>The activity was completed.</td>
</tr>
<tr>
<td></td>
<td>Session 6</td>
<td>23 minutes</td>
<td>The parent moved to the report questions.</td>
</tr>
<tr>
<td>Iteration 3</td>
<td>Session 7</td>
<td>34 minutes</td>
<td>The researcher told them they can stop.</td>
</tr>
<tr>
<td></td>
<td>Session 8</td>
<td>28 minutes</td>
<td>The parent moved to the report questions.</td>
</tr>
</tbody>
</table>
Table 10 continued

| Session 9 | 18 minutes | The parent asked if the child was done. |

**Frequency of conducting crowdsourced usability tests.** When the concept of crowdsourced usability testing was explained to the parents, all but one did not have concerns regarding the fact that the child is allowed to test products frequently. Parent 8 had concerns regarding the unlimited and uncontrolled use of technology for children. She said:

Well, kids now spend a lot of time using technology which I think it’s fine in the sense it’s what’s available now, but at the same time I think there are a lot of things that they miss out if we just let them be on those devices without limiting their time, because you want the kids to also make art or to read, and if you give a kid something he’s going to be playing around, they’re going to forget about all the other stuff. So, I try to balance that. So, I don’t know, maybe we can do it once or twice a month or something like that.

On the other hand, parents 6 and 7 said they would let their children use the website to test products every day. Some parents said it is up to the child. Parent 2 said, “I think this is up to her; she likes it, so I wouldn’t mind if it’s taking few minutes.” Parent 4 also had a similar answer. She said, “if she wants to, yeah, why not.” Parent 9 said that her child wants to be a “Youtuber,” and that she thinks he would love to record himself knowing that he is helping the instructional designers. She explained:

I’d probably say he can do it in the weekends, you know, like he could do it probably once or twice a week, and I would, you know, let him decide that
mostly, but I would only, like if he had to do it, it’d probably be once a week, and
then if he wanted to do it more, if he really liked it and stuff, I think I would, you
know, let him do it more, and I know him, he’d probably try to incorporate and
then just recording himself doing it and put it online, so.

Some have pointed to the positive side of testing products frequently. Parent 5
said, “I like the idea that she’s learning something and trying to get the hang of it by
herself, so that’s by itself to me is learning. So, I would go unlimited with this.”

Similarly, parent 3 said:

Maybe once every two or three days. Because if she’s free, I want to find some
activities for her to do while she’s free, those activities should be adding value to
her. I believe this one is one of these activities that would add value for her, she’s
going to learn from. So yeah, I’ll be encouraging her [to use it].

Parent 7 also had a similar answer:

I would let her do it every day. I think it’s cool, I definitely think it develops
critical thinking, and it teaches her new things, and it teaches her area that she
wants to practice and work on, and she likes challenges. So, “oh, I couldn’t do it,
I want to go back and try again,” so, I think this is loading her capacity.

**Benefit.** As explained earlier in Chapter 1 and Chapter 2, the undertaking of a
task in crowdsourcing websites always entails a mutual benefit for the crowd and the
crowdsourcer. Since the proposed approach in this study requires the child to frequently
test products and earn some kind of benefit, it was important to investigate the parents’
opinion regarding this issue, especially that all the children who participated in this study
were excited when they earned their gift cards. The parents were asked whether they
have any concerns regarding the monetary reward their children could earn each time they test a product, and whether there are any other kinds of benefits or rewards that are acceptable in such websites.

Two parents said that since this is a time-consuming activity, there should be a reward. Parent 4 thinks that parents will not be interested in participating unless there was something in return. She said:

It’s not too much work, but it’s consuming. If my child knew that he’s going to be rewarded, and that I will be rewarded, then probably I don’t mind spending an hour, you know, or an hour and a half or two.”

Parent 9 also explained:

It takes some investment of time. So, I think definitely you will have to give something in return, you know, some kind of reward maybe, you know, being able to access the game for some time, when it’s already finished or something.

When interviewed, none of the parents showed concerns about their children earning money at the end of usability testing sessions. Parents 2, 3, 7, and 9 compared the situation with a real-life job, seeing this as a learning opportunity that teaches the children the value of money and the concept of earning it when doing a task or a good thing. Parent 2 said:

I think this is a good thing because she will know the value of what she is doing, and it’s like a job for her, and part of it is fun because she likes to go through games and websites, and she’s on her laptop all the time, so it’s good to have something in return.

Parent 3 had a similar answer:
She has to feel like she’s taking money because she’s doing something, which our life is all about, really (laughs). We take money because we do jobs, so I want her to feel like she’s now, this is a kind of independency.

Similarly, when parent 7 was asked about her concerns regarding children earning monetary rewards, she said:

I think that is good because, well for two reasons, she’s at the point of thinking about money and how money works, and so I think it’s a great tool for her to say “okay,” to get her relationship with money, but on the other end, just knowing that doing good things, you know, there are positive outcomes for it, so.

Parent 9 said, “I think it’s life, it’s how we work, and we get paid for it,” and when she was asked if it is fine for a child, she said:

For a nine-year old, I think, or for any. Yeah, and he’s learning about doing something and earning money, it’s not like somebody is just giving it to him but doing nothing, so, yeah, so I think it’s fine, there’s nothing wrong with it.

Some parents emphasized the importance of the parent’s supervision. When parent 4 was asked about her concerns regarding the monetary rewards, she said, “because it’s under the supervision of the parent, it should be safe.” Similarly, when asked if she had any concerns, parent 7 said, “no, because I will make sure that she knows how to spend it.” Later she suggested, “maybe it’s better if you tell the child that it’s her money, because she made it, but it’s under the mother’s supervision.”

The parents were asked if they could think of any acceptable alternatives that the child can be rewarded with, other than the monetary reward. Parent 1 said that her child would be more interested if it is for money. Parent 5 said that “the reward is learning and
playing by itself.” Some of the other parents suggestions were: Free subscriptions or access to children’s websites, earning scores and using them to build a character, points to unlock new levels, and free toys. One parent jokingly suggested bitcoins, but then she said, “I think “good job” or “thank you” is fine.”

Finally, the parents were asked about their opinion in offering mini-payments, which sometimes might not exceed a dollar, for mini-tasks that the children can do in shorter time. This kind of rewards are used in some crowdsourcing services. The parents had different reactions. Some thought that it was a good idea. Parent 2 said, “I think the mini-tasks is a very good idea because it will help her be more focused, she will be like “OK, I’m looking at this thing and this is what they want me to do.” She did not have any concerns regarding the mini-payments. Parent 3 explained how this approach, in his opinion, might attract the child. He said:

It’s like a game “oh, if I finish this I’ll take ten cents, next one is 20 cents, oh, now how much do I have?” the more you take, maybe it will encourage children to do more work, it’s like a commission.

When asked if this is a good thing, he said, “yeah … I think they will do it faster and more efficient,” but from his opinion as a parent, he said it does not make a difference whether the reward was a whole payment for a whole task or a mini-payments for mini-tasks.

Parent 4 also compared this to a game. She said it makes it more interactive and engaging, but it depends on the child. She said, “if I think about my children, I will think of providing multiple payments for my second child it would be very exciting, but for her, it wouldn’t.” She suggested giving the child the choice whether to have the whole
payment at the end of the session or to do mini-task and earn mini-payments. Similarly, parent 5 said that mini-payments would be motivating and that she does not have any concerns about it. Parent 7 said that although she said that she would let her child use the usability testing website every day, it is going to be hard when there is homework to be done. She thinks that the mini-tasks will be a good idea for these days. She said, “I think that she would be very excited, putting nickels and dimes and a penny, she’s always excited about change, so I think she will be even more excited.”

On the contrary, some parents did not prefer the mini-payments idea. Parent 1 said that in mini-payments, the children might “do it only for the money, so you will not get an honest feedback.” She also said that she does not want her child to do it only for the money. Parent 6 said that the children “will not enjoy the website” when working on the mini-tasks, and that they “will just want to get the money,” but will not learn anything from the website they are testing. Parent 8 explained her concerns regarding the mini-payments:

I think it can get pretty addictive. I can see the child going like “ten more, and ten, and ten,” you know, I think. Because you’re supposedly you don’t do it just for the money. So, I wouldn’t want that to be the point they get across.

In general, there were no concerns regarding the regular payment that the children earn at the end of the session. Parents think that it is a good opportunity for the children to learn a real-life skill. However, there were different opinions regarding the mini-payments. Most of them believed that it is more engaging and motivating. Some perceived this as a positive point, while the others explained the negative side of it. Some parents pointed out that the choice whether to earn a whole reward or a mini-payment...
should be left to the children. Other alternatives to the monetary rewards were suggested, but all the children during the observation were excited when received the gift cards. This was also acknowledged by the parents during the interviews. Parent 9 said, “he’s really excited about that card, you know, he’d probably take mine too,” and parent 4 said, “she was really excited, and as I said, I’m not sure why, but probably the reward is why.”

**Other findings related to the parents’ concerns.** Other themes have been identified from the parents’ interviews. One of these themes is that the parents’ concerns regarding the children’s participation in crowdsourced usability testing is linked to the parents’ presence and the type of products to be tested. When parent 4 was asked about her concerns, she said, “I mean these are learning websites, right? So, I can’t think of any concerns. I don’t know, like I mean especially that I’m with her, so I don’t know if there is any.” Parent 8 also said, “as long as the parent is with the child all the time while they’re doing that, there shouldn’t be a problem.” Parent 3 had the same answer. He said, “as I’m with her, I don’t think there is a problem.” When parent 5 was asked if she has concerns regarding the content of the products, she said, “I guess it depends, since it’s educational, I would go %100 trust.”

Another point that is related to the parents’ concerns is whether the repeated testing of defective or incomplete products would negatively affect the child. Several parents believe that usability testing can have positive effects on the child. Parents 1 and 5 said that it makes children develop critical thinking. Parent 1 also added that it “probably gives [the children] some confidence that somebody is taking [their] opinion.” The same opinion was restated by parent 8. She said that usability testing can make the
children “more empowered because they are participating and have a word on something that adults make, and they can feel that they are important and that their opinions are being valued.” Parent 2 said that she thinks usability testing can be helpful for her child because it helps her know “what are the positive and negative things in the website … instead of just playing.” Both parents 3 and 5 mentioned that they believe that some of the advantages of participating in usability testing are discovering new things and learning how to evaluate them.

Even though most parents thought of crowdsourced usability testing as an experience that would impose positive effects on the children, some parents also had concerns regarding the repeated testing of defective products. Parent 7 said that if the children kept receiving negative feedback from the products they are testing, or if they kept not being able to complete the tasks because of problems in the products themselves, “they might internalize not meeting the game’s levels, I don’t know, there’s something maybe wrong, or that I’m not smart enough to do this.” Parent 9 also said that “some kids are super laid back and super confident in themselves, and some kids get really frustrated when they don’t know how to do something.” She thinks that it is better for the parent whose child is from the second type “who gets frustrated with themselves really easily,” it would be better to say, “you can’t use it that much, or that, you know, you should limit your use on this,” unless they are told beforehand that the products they are testing are defective so that they can discover the problems for the developers.

Summary

This chapter provided the findings from the design process of a remote usability testing tool, Kid Crowd, which was designed based on Hanna’s et al. (1997) guidelines
and the principles of cognitive load theory. Nine children and nine parents participated in the study. Each one of them participated in one of three iterations. The findings are summarized below. The number of usability problems reported by the parents in each iteration are shown in Table 11.

Table 11

*Number of Reported Usability Problems by Parents in Each Iteration*

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Number of Usability Problems Reported by Parents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iteration 1</td>
<td>4</td>
</tr>
<tr>
<td>Iteration 2</td>
<td>6</td>
</tr>
<tr>
<td>Iteration 3</td>
<td>18</td>
</tr>
</tbody>
</table>

**Children-related findings.** These findings are about the features of the remote usability testing tool that are related to children nine to ten years old, and they can be summarized as follows:

1. Some children found difficulties reading the font that was appropriate for most children.
2. Children showed positive reactions when they were told that their parents are going to be with them during the session.
3. Children seemed comfortable while they were being video recorded during the session, in which their faces were not showing on the screen.
4. During the session, children imitated what they saw in the video example of a usability testing session.

5. Children imitated their parents during the session in talking or staying silent, speaking loudly or softly, and in speaking in a specific language.

6. Some children were shy of telling their sincere opinion about the product.

7. When they were asked about their opinion in the website they tested, children were confused between that website and the usability testing website.

8. Children seemed bored when parents were slow in typing their feedback.

9. When parents discussed the answers with the children, the children’s answers were more detailed than when they entered their answers without parents’ presence or interference.

10. Children did not seem tired, and they did not ask for a break during the sessions, which lasted between 37 and 63 minutes.

11. Some children seemed anxious about the timer.

12. Children were excited and happy to earn the reward at the end of the session.

Parents-related findings. These are the findings about the features of the remote usability testing tool that are related to the parents or that could help the parents provide better usability testing reports with less workload. They can be summarized as follows:

1. Some parents could not determine if they were advanced or fluent in English language, and the number of years they lived in the country was not always a good indicator of their language skills.

2. Parents with slow internet connection could not complete the video instructions or had to watch it in parts.
3. Parents needed detailed instructions on how to set-up the testing room and equipment.

4. Having a short practice with a video example of a usability testing session seemed to help the parents in understanding their role and being prepared for the session.

5. Parents expressed their need to have the instructions in different forms: Written, and video instructions.
   a. It was found that some parents forgot the instructions that were in a part of the video that violated one of the principles of cognitive load theory.
   b. Parents preferred watching the instructions in shorter videos than watching one long video.
   c. Most parents did not read the written instructions attentively; highlighting the text and screenshots seemed to help.

6. Parents expressed their need to know more about how to use the product to be tested in order to be able to concentrate on providing feedback during the session rather than trying to figure out how to use it.

7. Some of the factors that might have affected the parents’ feedback are: The parent’s knowledge about the topic of the product to be tested, typing speed, language proficiency, how prepared the parent was before starting the session, the way the report questions were written, and how much the parent encouraged the child to talk during the session.

8. Some parents found it challenging to work on two devices at the same time, but they think that this is normal since it was their first time trying such an experience.
9. Parents worked more smoothly when they were prepared for what is coming next, and when they knew who was supposed to do what and when.

10. Parents helped the children provide more sincere answers when they felt that they were not saying what they were thinking, even though some of them forgot to watch the facial expressions and the body language of the children.

11. Most parents found the observer’s role and the technical troubleshooter’s role to be the most challenging roles in the session.

**Findings about the parents’ concerns.** Findings about the parents’ concerns regarding children’s participation in crowdsourced usability testing can be divided into four sections:

1. Parents’ concerns regarding safety and privacy:
   a. The parents’ approval for their children to participate in a crowdsourced usability testing website without the help of the parent depends on how trusted and safe the website is.
   b. The parents’ concerns regarding safety and privacy depend on whether they are with the children while using the website or not.
   c. Parents are mostly concerned about two issues: The privacy of identity, which is mainly represented by the parents’ concerns regarding videotaping their children, and the appropriateness of content, which is represented by the parents’ concerns regarding the appropriateness of the products to be tested in terms of age and culture.

2. Parents’ concerns regarding how long and how often their children can participate in usability testing:
a. Most parents perceived their children’s participation in usability testing as a positive experience.

b. Some parents believe that a usability testing session should not be longer than an hour for a comfortable experience for the child and the parent, while the others believe it should not be longer than fifteen minutes.

c. Most parents are against setting a timer for the usability test. They believe it has negative effects on the child’s and the parent’s performance.

d. Most parents are willing to allow their children to frequently participate in crowdsourced usability testing, while some emphasize the importance of controlling the children’s participation so that it does not become addictive.

3. Parents’ concerns regarding the reward that the children obtain from participating in crowdsourced usability testing:

a. Parents do not have concerns regarding the regular monetary rewards as long as it is under the parent’s supervision.

b. While some parents believe that the mini-payments that are paid for mini-tasks can engage the children into doing more tasks, other parents are afraid that this type of reward could be addictive, or that it could make the children care about the money more than the learning part of the process.

c. Other types of rewards have been suggested as alternatives.

4. Other concerns:
a. Some parents are concerned about the children’s feelings when they are constantly being provided with negative feedback from testing defective products.
Chapter 5: Discussion and Conclusion

The aim of this study was to provide a set of guidelines for the remote usability testing of children’s interactive products. The design and development research methodology was used to answer the following research questions:

1. What can be learned about the remote usability testing guidelines in terms of:
   a. the features and characteristics of the usability testing tool that are related to the children nine to ten years old?
   b. improving the role of parents in such a way that can reduce the workload of parents while maintaining a good quality of usability testing results?

2. What concerns do parents have regarding their children’s participation in crowdsourced usability testing?

A prototype of a crowdsourced usability testing tool, Kid Crowd, was designed based on Hanna’s et al. (1997) guidelines and the principles of cognitive load theory. Nine parents and nine children participated in one of three iterations in this study. Data collection was conducted using observations, fieldnotes, and interviews.

In this chapter, a discussion of the findings and recommendations for practitioners are provided in light of previous research in order to propose tentative guidelines for the design of such tools. The guidelines are derived from what went well and what did not, which was noticed during the sessions and discussed with the parents during the interviews. The guidelines are also informed by the parents’ concerns regarding their children’s participation in crowdsourced usability testing. In addition, this chapter provides recommendations for future research.
Children-Related Findings

This section provides a discussion of what did and did not go well with the children nine to ten years old during the remote usability testing session.

**Parent’s presence.** The children were comfortable and showed positive reactions when they were told that their parents are going to stay with them during the session. The parents’ presence also affected the children’s feedback. Children who discussed their answers with their parents provided helpful feedback with more details than children who answered the report’s questions alone. Most parents also believe that their children cannot provide the same feedback if they were alone. The parents’ presence also helped in putting the children’s responses into context. When the children were shy to say the truth about how they feel about the product they tested, the parents indicated that in their reports and helped their children be sincerer. These findings are in line with those of Markopoulos et al. (2008). They indicated that the advantage of testing at home with the help of the parent is that it provides a familiar environment for the child who feels more comfortable answering his or her parents’ questions, especially that the testing situation is similar to real life situations in which the product is usually introduced by the parent. Khanum & Trivedi (2013a) also found that children are more relaxed in the field. Markopoulos et al. (2008) also stated that the parent’s presence adds to the value of the testing results since they know the child and therefore can put the child’s responses or performance into context.

**Unobtrusive videotaping.** Children were also comfortable in front of the camera. They were told in the assent form that they will be videotaped during the session in order for the researcher to watch the video later and take notes. The software that was
used to record the session used the laptop’s built in camera, and it only showed the face of the child at the beginning of the session. During the session, the recording was unobtrusive. Even the children who were usually shy in front of the camera did not seem to be shy or distracted during the session. These findings are consistent with Hanna’s et al. (1997) guidelines, which suggest that the testing equipment should be selected and placed wisely and unobtrusively in the lab to capture the optimum sound and picture in a least-distracting manner.

**Break.** The sessions in this study did not include a break for the children, but the parents were instructed to offer a short break if the session was longer than 45 minutes. However, the parents did not offer a break because none of the children seemed tired or asked for a break. The parents said that there was no need for a break. Most of them said that they do not prefer having a break in the middle of the session, and that it might actually distract the children and interrupt the consistency of the session. These findings run counter to Hanna’s et al. (1997) who stated that the children in the lab need to have a break if the session is longer than 45 minutes. Since the sessions in this study were conducted at home with the presence of the parent, the experience was more natural and similar to the children’s everyday use of computers. Therefore, there was no need to set a specific break period unless the child needed to have a break. In general, the sessions went well without offering any breaks for the children.

**Typing the report.** As mentioned earlier, when the children were encouraged to discuss their answers with their parents before typing them, they provided more detailed and useful usability testing reports. However, when they were asked to type the report without the parents’ interference, their reports were brief and sometimes not helpful in
terms of informing the designer of the products’ usability problems. On the other hand, when the parents were asked to type the children’s answers, some children were shy or hesitant to say their true opinion about the product. The parents suggested to let the children type the report alone in order to obtain more honest answers without having to face “guilt and embarrassment” in front of the parent. This is consistent with what Rubin and Chisnell (2008) stated about self-reporting usability testing. They said that these kinds of questionnaires need to be conducted in an unobtrusive way to avoid leading the users to alter their behavior while performing the tasks (Rubin & Chisnell, 2008).

As the children’s reaction might have been affected by the parent’s presence, it could also be the effect of the researcher’s presence that led the children to be shy. Markopoulos et al. (2008) stated that some children feel shy or uncomfortable around strangers, especially when they think that they are being tested and not the product, which affects their ability to verbalize their thoughts, or it leads them to mention problems that do not exist just to please the observer. Interestingly, what happened in this study was the opposite, even though either situation would have caused the children’s reports to be unrepresentative of what they think; some children tried to pretend that they did not find any problems in the product to please the researcher, who they thought created the product they were testing. Once they knew that the researcher did not design the product, they felt more comfortable and started saying what they really felt.

**Verbalizing the children’s thoughts.** It was noticed that the parents had a strong influence on the children’s verbalization. The children copied their parents in verbalizing their thoughts. In other words, their verbalization depended on the parents’ attitude in terms of the volume, the language they used, and the amount of verbalization. The
children of the parents who whispered were whispering during the session. The children of the parents who spoke in a specific language during the session also replied to them using that language. Similarly, the children of the parents who stayed silent were found to be silent and did not try to think aloud except for those who watched an example of a child thinking aloud at the beginning of the session; the children in this case imitated the child in the video. These findings can be linked to the social learning theory (Bandura & Walters, 1977), also known as social modeling, which states that people learn from others through observation and imitation.

These findings also add to those of Khanum and Trivedi’s (2013b) who found that the children provided better verbalization in the field. The current study suggests that if the parents are monitoring the session in the field, then the children’s verbalization depends to some extent on the parents’ verbalization. Also, as Markopoulos et al. (2008) suggested, parents should encourage the children to verbalize their thoughts through a dialogical approach, rather than depending on the thinking aloud strategy, in order to offer more guidance and to reduce the cognitive load imposed on the child. They found that the children are more likely to not think aloud when facing a problem due to the high cognitive load at these moments. In the current study, parents who tried to have a conversation with the children during the session helped them verbalize their thoughts and were able to guide them through the usability test. For those who did not, their children were silent most of the time, and they were not able to guide them in order to discover more problems.

It was also noticed that the children who watched a short video example of a child thinking aloud right before the session did think aloud even though their parents were
silent. Nielsen stated that a great way that helps when using this approach is “showing the test user a short video of somebody else thinking aloud” (2014, para. 3). He said that “demonstrating something by a concrete example gets the point across better than most abstract explanations” (2014, para. 3). However, when the example provided in this study included a product that is similar to the one to be tested, the children started imitating the actions and the answers of the child in the example. This problem was solved when the example was changed to a product that is totally different than the one the children were supposed to test. Also, when the child in the example provided some negative and positive feedback on the product she was testing, the children were found to provide negative feedback without being shy or hesitant. These findings are also consistent with Nielsen’s who said that the video “should show a different user interface” in order to not bias the users, and that it should include the key points that the instructional designer need to learn from the test (2014, para. 6).

**Timer.** Setting a time for the usability testing session was not preferable neither by the children nor the parents. Some children were anxious and upset about having the timer. Parents also did not see a point of specifying a limited time for the session. Some also mentioned that the timer affected their feedback and caused some confusion because they were not sure whether to wait until the time was up and then type their observations, or whether they were supposed to start typing their observations while the timer is counting the time. These findings are in line with those that mentioned that the quality of the feedback in a usability testing session is affected by the length of the testing session (Hwang & Salvendy 2010) and the time pressure (Mäntylä, 2013).
Others. It was also found that the designer needs to think of the unexpected. Unlike in controllable environments like usability testing labs, when a usability testing session is conducted at the participants’ home, things like the internet connection, the parent and child position, and the recording equipment can affect the quality of the participants’ feedback. In this study, the slow internet connection in some sessions affected the streaming of the instructions video. The children and the parents in these sessions were not able to watch the videos normally. This could have been fixed if the videos were available in different quality settings. Also, another example is when one of the children sat behind a light source, which made it hard to observe her face expressions in the video recording.

As discussed in Chapter 2, the traditional usability testing approach requires strict controlling procedures that are more like those used in experimental research (Barnum, 2011; Rubin & Chisnell, 2008). However, according to Rubin and Chisnell (2008), it is not only difficult to meet the conditions of the experimental design, but it is also unnecessary since the goal of experimental design is to test hypotheses while the goal of usability testing is to improve specific products. Barnum (2011) also stated that it is illogical to waste money and time in recruiting a large number of randomly selected participants and conduct the test under certain conditions in order to achieve a goal that can be achieved in an informal and more flexible way. Testing in the field, on the other hand, provides higher ecological validity since it simulates the actual use of the product and facilitates the detection of real usability problems (Markopoulos et al., 2008). However, based on the findings of the current study, the control that is offered in the traditional approaches might not be crucial for the improvement of the product per se, but
it could be helpful in situations like the ones described above. In other words, it can be helpful for the situations that cannot always be expected because of the variability of the settings or the variability of the parents’ experience with the technical issues, for example, which might hinder the quality of the feedback.

There were other unexpected findings that are not related to the physical setting. For example, one child found difficulties reading the font that was appropriate for all the other children, and that was approved by an elementary school teacher. The parent suggested having flexible font size that can be controlled by the child. Another example is that when the children were asked about their opinion in the website they tested, they were confused between the usability testing website and the product that they were supposed to test; in more than one session, the children were noticed asking their parents whether the questions were about Kid Crowd or the product they were testing. The problem was solved by placing a webpage, between the instructions in the usability testing website and the website to be tested, that included a note that clearly states that the children will move to the website that needs to be tested. The point is that it is difficult to expect everything that is going to happen in a remote usability test with children, and with the absence of an experienced test moderator, pilot testing of the remote usability testing tool becomes crucial. Markopoulou et al., (2008) and Schade (2015) also emphasized the importance of pilot testing before conducting the actual usability testing sessions. The next section discussed the findings that are related to the second research question.
Parents-Related Findings

This section provides a discussion of what went well and what did not during the remote usability testing session with the parents in terms of being able to provide good usability testing results with reduced workload.

**Being prepared.** One of the findings in this study is that when the parents were prepared for their role, and when they expected what is coming before beginning the session, the whole procedure went more smoothly with less confusion, which helped in collecting better feedback from the parents. Simply informing the parent about the objectives and their role was not enough. One way that was found to be helpful is to allow the parent to practice his or her main role in the session by watching a short example of a parent and a child performing a usability test while entering his or her observations. When the video ends, the parent could then compare his or her report with the report of the parent in the example.

As mentioned earlier, Nielsen emphasized the importance of demonstrating thinking aloud by providing a video example of a person thinking aloud (2014). Using video examples was not only helpful in demonstrating thinking aloud, but it was also helpful in demonstrating the role of the parent as a test administrator and an observer. However, the parents said that the example alone would not have made a difference without the part in which they compared their reports with the report in the example. In terms of the cognitive load theory, this could be thought of as an instance of completion problems, which is supposed to “decrease extraneous cognitive load, facilitate the construction of schemas, and lead to better transfer performance” (Sweller et al., 1998, p. 277).
Parents also worked more smoothly when they knew who was supposed to do what and when. In this kind of usability testing, the session can easily become complicated because there are two people who should do specific things at specific times and in a specific order on two different devices. It can become confusing if it is not clear who should do what and when. It was found that dividing the instructions visually into steps was helpful. For example, in this study, the parents’ screen had three steps. Once one of them is clicked, the details of the other two become hidden. The child’s device had notes to help the child remind his or her parent of the next step once he or she is done with the current task. This was added when it was noticed that sometimes the parent would be looking at the child’s screen and forgets to read the instructions on his or her own screen. In other words, this redundancy of the instructions was intentionally designed to cover for the split attention effect that happens from looking at two different sources of information in order to understand the content (Sweller et al., 1998).

Parents said that the technical troubleshooter’s role and the observer’s role were the most challenging roles in the session. Therefore, designers need to pay special attention to the design of the instructions of these two roles. Also, more details need to be given in the set-up instructions because there could be situations that would not occur in the lab, or that are intuitive to the more experienced test administrators, but not for the parents. For example, recording a session behind a light source or in a room with a loud air conditioner. These situations could affect the quality of the recordings, but the parents might not be aware of. Therefore, the literature strongly recommends conducting a pilot study for the usability test (Markopoulos et al., 2008; Schade, 2015). Pilot testing can help in creating more comprehensive instructions for the parents.
Speaking about a usability testing method that requires the parent to take the roles usually taken by the test administrator, the observer, the technical troubleshooter, and the diarist, Markopoulos et al. (2008) emphasized the need to train the parents on their role and to find ways that reduce their workload. As described above, a short practice in the form of a usability testing completion problem was found to be helpful in this study. However, since this is a crowdsourced usability testing method, the participants are supposed to repeat the process several times. The parents in this study believe that the process will be easier the more they perform the test. This is compatible with the fading guidance effect in the cognitive load theory. According to Sweller et al. (2011), as the learner gains experience, the elements that are necessary to learn a specific knowledge component are integrated into schemas, which reduces the load on the working memory, which requires the designers to fade the guidance as experience increases. However, the current study only investigated the first participation by the parents and their children in the crowdsourced usability testing tool.

Finally, it was found that the parents can be more prepared for the testing session when they understand what the product to be tested is about. Parents expressed their need to know more about how to use the product in order to be able to concentrate on providing feedback during the session rather than trying to figure out how to use the product. This could be done by providing short videos that explain the product to the parent. Having limited background about the field of the product was found to affect the parents’ ability to encourage verbalization by having a dialog with the children during the session, and thus it affected their ability to discover more usability problems. Markopoulos et al. (2008) stated that having a dialog with the child during a usability test
“allows more social interaction and more guidance to offer” (Markopoulos et al., 2008, Chapter 10, Section 1, para. 3). This is easier when the person moderating the usability test is the designer of the product. But when the test moderator is the parent, his or her lack of knowledge about the product and its field could affect his ability to encourage thinking aloud through a dialogical approach, his ability to guide the child during the session, and consequently his ability to discover more usability problems.

**Instructions.** One of the factors that seemed to affect the parents’ understanding of their role, and thus their performance and feedback in the usability test, is how clear the instructions were. Parents expressed their need to be able to choose the modality of instructions as some of them prefer written while the others prefer video instructions. This need was expressed when some parents commented on the video instructions that it was too long and has a lot of information. However, when the videos were divided into shorter ones, and the written instructions were added, all of the participants chose to watch the video version rather than reading the instructions, even those who later said that they prefer reading over watching. They said that if the videos were not short and straightforward, they would have checked the written instructions.

Although the content was the same, and the only thing that was changed in the first version of the video was that it was divided into three shorter ones, some parents described the instructions in the long video as “very long” and has a lot of information to “comprehend and process at the same time,” while they described the shorter videos as “easy” and “straightforward.” This might be linked to the fact that the capacity of working memory was found to be limited to 7±2 items of information at a time (Miller, 1956). In other words, the human’s short-term memory can store between five to nine
objects at a time. Therefore, the design of the instructional materials can either overwhelm or facilitate the job of the working memory (Sweller, 2008).

When the parents had no choice but to read the instructions, which sometimes included screenshots, most of them scanned the texts. They did not seem to read the instructions attentively, which was also clear from their performance during the observations and from their answers during the interviews. What helped in this case was to divide the written instructions into blocks; to highlight the important statements either by making them bold, or changing their color to make them stand out; and to highlight the important parts in the images and instructions using red arrows and visual symbols, which were added to the video instructions as well. This technique is known as “cueing,” which is highlighting the important parts of the content using non-content objects (Meyer, 1975). Cueing was found to enhance comprehension and/or transfer when used in animated videos (de Koning, Tabbers, Rikers, & Paas, 2007) images (Tabbers, Martens, & van Merriënboer, 2004), and texts (Lorch & Lorch, 1996).

Although it was not possible in this study to measure the cognitive load, and therefore, the researcher has to be cautious when drawing conclusions, but an incident that is related to the split attention effect was noticed in a number of sessions. Previous research indicates that the “simultaneous presentation of the narrative and visual portions of an instructional movie” results in better recall and problem-solving transfer (Mayer & Anderson, 1991, p. 490). One part of the parents’ instructions video unintentionally had the split attention effect. Each piece of instructions in the video was accompanied with a simple animation, except one that the researcher thought of as part of the previous piece of instructions, and therefore they both had the same animation. In other words, this was
the only part in the video that did not have its own animation. It was found that during the sessions, several parents remembered the first piece of instructions that accompanied that animation and forgot the second part; they only remembered it when they were reminded about it in the interviews. Therefore, the designers need to be cautious when designing the instructions following the cognitive load theory.

The last finding that is related to the instructions is the modality of the instructions and the problem of slow internet connection. This problem in some sessions caused the videos to pause frequently. Some parents who faced this problem seemed bored, chose not to complete the instructions, and said that they could not focus when they were watching the instructions. Therefore, it is important for these situations to provide a different form of instructions, and to provide the option of selecting the quality of the video in order to solve the problem of slow streaming.

**Challenging roles and tasks.** Markopoulou et al. (2008) stated that in the usability testing method that is conducted at home without the presence of a test moderator, the parent’s role is maximized, and therefore, there should be ways to reduce the workload on the parent. The observer role was identified as the most challenging role for the parents in the remote usability tests. This role starts when the child begins using the product, and it includes reading the tasks for the child, watching him or her performing the tasks, identifying usability problems, and entering them in the report form.

The parents identified several reasons why the observer role was the most challenging one. First, working on two devices at the same time was difficult for the parents who are not experienced with those kinds of tasks. This could be linked to the
split attention effect. Symonds (2011) discussed this problem in another usability testing method. She referred to the problem of moving between the survey’s page and the website under investigation as the users’ distributed attention, which could affect the quality of the feedback provided by the users. However, the participants stated that it is normal to experience that in the first time they perform the test, and that they expect themselves to overcome this problem once they understand and experience the whole process more than once. This can be linked to the findings that indicate that one of the factors that affect the quality of the feedback provided in a usability test is the participants’ expertise in usability testing. The more experienced the participants are, the more usability problems they can discover (Mäntylä, 2013; Hwang & Salvendy, 2010).

According to the cognitive load theory, the size and number of schemas that are stored in the long-term memory, which reduces the load on the working memory, determine the level of expertise of a person in a specific task (Sweller, 1988). Two cognitive load theory effects are related to this point: The expertise reversal effect, which suggests that as the level of prior knowledge and expertise increases, instructional procedures that are effective with novice learners might become a burden for experienced learners (Sweller, 2008); and the guidance fading effect, which suggests that because previously hard to comprehend information becomes easier for more experienced learners, the provided guidance should gradually fades as the learner’s expertise increases (Sweller, Ayres, & Kalyuga, 2011).

The parents also identified another factor that contributed in making the observer role more challenging. In this study, eight out of nine parents had no experience with coding or coding games, which was the field of the website being tested by their children.
This lack of knowledge about the topic of the product under investigation added one more mental task for the parents to do during the observation time, which is trying to learn how to use the product, and trying to understand whether their children were using it correctly in order to discover the usability problems.

Markopoulos et al. (2008) emphasized the importance of training the parents who take all the roles in a usability test. A short example of a usability testing session with the ability to practice taking notes while watching the video was found to be helpful in this study. In addition, based on the findings of the current study, parents also need to be introduced to the type of product that will be tested by their children. A short video that includes the goals of the product and how to use it could be helpful. Furthermore, since this is crowdsourced usability testing, which means that the parents are expected to repeat the experience several times, this could mean that the more they perform the experience, the easier it would become for them. This piece of data could be added to the parents’ profiles in order to control the amount of guidance provided based on the guidance fading effect of the cognitive load theory (Sweller, Ayres, & Kalyuga, 2011).

**Usability testing report.** The quality of the parents’ feedback in the current study was found to be affected by several factors. The more obvious factors are the typing skills and the English language proficiency of the parents. It was noticed that the parents who were able to type fast without looking at their screens provided longer reports and were able to keep observing their children’s screens even while typing their observations. Parents who were slower and who had to look at their screens while typing their observations were not able to concentrate on the children’s screens all the time, which led them sometimes to miss some usability problems. Parents with weaker English
language skills were also found to provide less helpful usability reports. Litman et al., (2014) also found that because crowdsourcing comprises people from different countries, it is a challenge for some of them to understand the instructions of the required task, which sometimes affects the quality of the feedback and results in irrelevant work.

One of the findings related to the parents’ language was that some parents were not sure which English language level to choose when they were filling their profiles in the website. Also, the number of years lived in a country is not always an indicator of how good the participants are in that country’s language. It might be more helpful to let the parents and the children take a language test that can help determine their language level. This can also be done with the typing speed. This kind of information can be helpful in recruitment, and can provide a convenient way of dividing the users into groups based on their attributes (Sherief et al., 2014).

One of the other factors that affected the quality of the usability testing reports provided by the parents is whether the parent encouraged the child to think aloud through a dialogical approach or whether they stayed silent. It was found that the parents who talked with their children during the session were able to capture more details, and they were also able to guide their children to try more paths in the product. This is consistent with Markopoulos’s et al. (2008) statement, which indicated that usability testing with children require “a more dialogical demeanor by the administrator, which allows more social interaction and more guidance to offer” rather than only depending on the think aloud approach (Markopoulos et al., 2008, Chapter 10, Section 1, para. 3). However, in this study, this factor was affected by how much the parents understood the product the children were testing, as discussed in the previous section. Parents’ who did not
understand the product stated that they did not ask about the things they did not understand, and they also stated that they were not able to “guide” their children and “support” them during the session. Nonetheless, it was noticed that some parents, who had no experience or knowledge about coding, but they followed the dialogical approach with their children during the session, it was noticed that they provided equally or more detailed reports than the parents who had coding experience but did not follow that approach.

Another factor that affected the quality of the reports provided by the parents is the way the questions of the report were written. This finding is in line with the researchers who found that “fine-tuning” the wording of the questions used in the crowdsourced tasks has a positive effect on the data quality since it reduces misunderstanding and confusion (Sarasua et al., 2012). For example, in the current study, parents stated that one of the reasons they did not provide more details is that the text boxes in which they were supposed to type their answers were labeled with the word “notes,” which made them feel that they only need to provide brief notes.

One observation that was noticed in this study is that the parents’ reports seemed to reflect the children’s true reactions, even though some of them forgot to watch the facial expressions and the body language of the children. They said that as parents, they can naturally tell how their children are feeling without intentionally looking at their faces or bodies. This is broadly in line with the literature, which indicates that the presence of the parents during the usability test adds to the value of the testing results since they know the child and therefore can put the child’s responses or performance into context (Markopoulos et al., 2008).
Finally, one of the strategies that was intended to be used to improve the quality of the parents’ feedback was asking the parents to review their report, adding any missing details, before submitting it. This approach was derived from one of the findings about how to increase the quality of the participants’ feedback in crowdsourcing environments. Goodman et al. (2013) indicated that one of the methods that proved to be effective is to simply ask the participants not to cheat, in terms of trying to complete the tasks as fast as they can in order to earn more rewards. However, simply asking the parents to review their answers in order to provide better feedback was not effective. It was found that almost none of the parents actually reviewed their answers before submitting them.

**Distractions.** In some of the sessions in the current study the parents were frequently interrupted by their other children in the house, or they could hear their noises during the usability test. However, this did not seem to distract the parents or the children performing the test; they did not seem to lose focus, and the reports they provided were representative of the problems that were faced during the session. According to Hanna et al. (1997), distractions by the siblings of the child can be controlled during the test at a lab; they suggest that siblings should not be allowed in the testing room. However, Khanum and Trivedi (2013a) found that interruptions during the test in the field setting did not influence the children’s performance, and that there was no significant difference between the problems detected in both settings.

**Findings on the Parents’ Concerns**

It was found that, in general, the parents do not have concerns about their children’s participation in a crowdsourced usability testing as long as they are with the children all the time during their participation. The parents’ approval for the children to
participate alone without the presence of the parent depends on the parents’ assurance that there are no safety or privacy issues. The parents were also asked about their concerns regarding the length and the frequency of their children’s participation, and about the monetary rewards that their children can earn from such participation.

**Safety and privacy.** Two safety and privacy issues were found to concern the parents: The privacy of identity and the appropriateness of the content. The parents’ concerns about the privacy of identity is mainly represented by their refusal of the idea of videotaping their children’s faces, and posting the recordings online to anonymous people. They believe that it is not essential since they are supposed to be present during the sessions, which means that they can substitute this type of recordings by watching and reporting the children’s face expressions and body language. Some parents indicated that if the usability testing website is trusted and authorized by a responsible authority, they would not mind recording their children’s faces. Some suggested that the website should provide a disclaimer or a privacy policy for the parents to understand what is going to happen to the recordings, and how they are going to be used. This only applies to the videotaping of the children’s faces. The parents did not express any concerns regarding the other types of recordings, such as the audio and the screen recordings.

The parents’ concerns are in line with the ethical considerations discussed in the literature. Nielsen (1993b) assured that the confidentiality of the test results and the privacy of the users’ identities should be maintained in usability testing in general, not only with children users. Markopoulos et al. (2008) emphasized the importance of informing the children that they will be photographed or video/audio recorded. They stated that permissions should be collected not only from the parent, but also from the
child. Sova and Nielsen (2008) also explained that sharing video recordings outside the usability testing team is considered unethical. Permission has to be collected if the video recording is going to be shared with any person outside the team (Nielsen, 1993b). Markopoulos et al. (2008) clarified that this rule does not only apply on video recordings, but also on audio recordings because children can be identified from their voices (2008).

One of the parents who accepted that her child be videotaped in crowdsourced usability testing website said that this is important for the designer to make sure that the child is indeed the person who tested the product. In fact, cheating is one of the identified crowdsourcing problems in the literature. This is because the crowd can earn monetary rewards for performing tasks or jobs in crowdsourcing websites. Therefore, some were found to cheat in different ways in order to earn the reward even without sincerely completing the tasks (Goodman et al., 2013; Ipeirotis et al., 2010; Rubin & Chisnell, 2008; Shah & Zhou, 2014). Several strategies were suggested and tried in the literature depending on the type of the task and the crowdsourcing platform. In crowdsourced usability testing with children, it is expected that the participants would pretend that they are children in order to collect the rewards. Videotaping could be one solution to solve this problem. However, the parents’ suggestions about the disclaimer and the authorization of the website need to be considered when designing such platforms.

The second issue that concerns the parents regarding safety and privacy is the appropriateness of the content. The parents are concerned about the content that the children can be exposed to in crowdsourced usability testing websites. Markopoulos et al. (2008) discussed this saying that there are two issues in usability testing sessions that require the children to use web content; these are the possibility for children to be
exposed to inappropriate content, and the possibility that they share personal information (Markopoulos et al., 2008). Inappropriate content usually describes the content that is inappropriate for a specific age group. However, in this study, when the parents were told that the idea of Kid Crowd is that none of the products to be tested can be posted on the website unless they meet certain standards of online safety and privacy for children, one parent said that it is not only about being age-appropriate. She said that a product can be appropriate in terms of age, but not culturally appropriate. Since crowdsourced usability testing is directed to a more diverse population that can include people from different backgrounds and cultures, there should be a way to make the parent check the product to be tested before choosing it. One parent also suggested having the ability to select a group of products that can be viewed in their account for testing after making sure that they are appropriate for the children’s age and culture.

One of the findings that can be classified under the safety of the children and the appropriateness of content is the parents’ concerns regarding the children’s repeated exposure to negative feedback from the products they are testing. They are concerned that this might impose negative psychological effects on the child. Also, because the products are supposed to be incomplete, this means that the children might not be able to complete the usability testing tasks because of the usability problems in the product itself. However, parents are concerned that this might lead them to think that they are not smart enough to complete the tasks successfully. Therefore, it should be made clear for the children that there are problems in the product they are testing, and that their job is to discover these problems.
**Length and frequency.** The parents were asked about how long they believe a usability testing session should last for, and how often they would let their children test products in a crowdsourced usability testing website. It was found that the parents prefer the more flexible approach where there is no specific time for the session, but most of them suggested that it should not exceed an hour. Some parents were more conservative and chose shorter durations. Some of them expressed their concerned regarding the overuse of technology for children.

One parent was concerned about children becoming addictive to participating in crowdsourced usability testing; she emphasized the importance of the parents’ control and supervision in order to avoid these situations, and in order to guide the children into spending their time doing things that are different than using computers and mobile devices. On the other hand, most parents perceived the experience as one that would benefit the children on the long run. Some of the benefits the parents believe their children could earn from participating in usability testing in general, and in crowdsourced usability testing in specific, include: Learning and exploring different topics and fields, developing critical thinking, feeling empowered and self-confident, building an experience and a “relationship” with money, and developing empathy towards others by helping the designers improve their products, and by contributing in providing better learning products for other children.

This finding was surprising because most of the literature on usability testing with children talks about the challenges and considerations of conducting this kind of usability testing (Barendregt et al., 2007; Byerly, 2007; Druin, 2002; Hanna et al., 1997; Khanum & Trivedi, 2012; Markopoulos et al., 2008; Markopoulos & Bekker, 2003; Rounding et
al., 2013), and the benefits of using the help of children users in terms of the results of the usability tests (Barendregt et al., 2007; Druin, 2002; Hanna et al., 1997; Markopoulos et al., 2008; Massion & Nam, 1998). The researcher did not find studies that talk about the benefits of usability testing on the children users. This is probably because in the traditional approaches of usability testing, the children are usually invited to test one product, or sometimes compare two products. However, in crowdsourced usability testing, the children and the parents would have accounts in the website in which they can sign in frequently and pick the products they want to test. They could be also invited through the platform based on their profile information. Being an experience that could be frequently conducted by the children, usability testing was viewed by the parents from a different angle that affected how they perceived its benefits and their concerns about it.

**Benefits and rewards.** One of the things that several researchers have investigated in previous research is the monetary reward that is given to the crowd after completing a task or a job in a crowdsourcing platform (Litman et al., 2014; Liu et al., 2012; Sherief et al., 2014; Zogaj et al., 2014). However, these studies were concerned about the adult crowd. Also, all the crowdsourcing platforms that were reviewed in this study were found to have a condition that does not allow children to participate in their services or collect the reward. To the best of our knowledge, the current study is the first study that suggests the use of crowdsourcing in usability testing with children test users. Therefore, it was important to investigate the parents’ concerns regarding the monetary rewards that could be earned by their children, and the other alternatives that could be offered in such platforms.
None of the parents had concerns regarding the regular monetary reward paid for the children after participating in crowdsourced usability testing if it is under the parents’ supervision. Some found it as a chance for the children to learn a real-life concept, which is earning money for doing a task or a “good thing.” Although most of the children accepted to participate in the usability test even without knowing about the monetary reward, they were all excited to earn the gift card at the end of the session. Some parents emphasized the importance for the children to participate in crowdsourced usability testing without money being the goal for such participation. Other parents said that the type of reward affects their decision whether to spend some time with their children to participate or not, especially that it takes some investment of time. When asked about the alternatives that could be offered other than monetary rewards, some said that the reward should appeal to the child. Some of the suggested options are: Free subscriptions in educational products or games, and scores that can be used to build a character, to unlock levels, or to buy things in the usability testing website itself.

Finally, when the parents were asked about their opinion in mini- or micro-payment, which is known to be used in some crowdsourcing websites as mentioned in Chapter 2, parents had different opinions. Although most of the parents indicated that this option can be a motivator for the children to use the website more frequently, some saw this from a positive angle, and others saw it from a negative one. For the parents who were with this option, they said that they believe it would appeal to the children because it seems like a game. It could be also useful for children who are busy with homework. For example, they could do a mini-task and earn a mini-payment for it, instead of spending longer time on longer tasks. Also, parents said that they think it
would appeal to the children because they would like to collect money and seeing it grow. On the other hand, the parents who were against mini-payments said that the children could become easily addictive to it; all what they would care about is the money part of the experience; it would limit their chances of learning from the products they are testing, and they might give insincere answers if all what they care about is collecting the money. The last point was noticed as a problem in crowdsourcing platforms, where some the people in the crowd were noticed cheating to earn as much money as they could in the shortest time possible (Goodman et al., 2013; Ipeirotis et al., 2010; Rubin & Chisnell, 2008; Shah & Zhou, 2014). However, these findings were all derived from the parents’ opinions not the children’s. To determine which type of payments would be more appealing and appropriate for children, the children themselves need to be asked those questions.

**Recommendations for Practitioners**

The purpose of this study was to provide a list of guidelines to help the practitioners design a remote usability testing tool for testing children’s interactive products. The following is a tentative set of guidelines, which represent the lessons learned from the iterative design of a crowdsourced usability testing tool, Kid Crowd. It is important to mention that these guidelines and recommendations will not become valid until they have been validated by future research. Some of these guidelines are specifically intended for crowdsourced usability testing tools, while the rest can be applied on remote usability testing tools in general.

1. Recommendations that are related to the children
a. If videotaping is necessary, it should be designed in an unobtrusive way so it does not distract the child.
b. Regarding the children’s instructions, it is recommended that the font size of the text to be large enough, or to allow the children to control the size of the text.
c. Watching a short example of verbalization in a usability testing session is recommended for the children. The video should include testing a product that is different than the product to be tested by the child; it should show different tasks than the ones in the current test; and it should include both positive and negative feedback.
d. The child should be able to differentiate between the usability testing website and the product he or she is testing. A page with a note that separates the usability testing website from the product can be helpful.
e. Setting a specific time for the session is not recommended. If necessary, it is better not to tell the child about it.
f. In case the children are better in the language of the report and faster in typing than their parents, it is recommended that they type their own answers in their part of the report after discussing them with their parents.
g. A short break should not be offered unless the child needs it.
h. Each remote usability testing tool is different in its details; it is important to conduct a pilot test before actually using the remote usability testing tool in order to avoid any problems that could affect the child’s feedback.

2. Recommendations that are related to the parents
a. Parents should be present to help in moderating the test and supporting their children.

b. The language proficiency and the typing speed of the participants should be tested and taken into consideration in the recruitment process. The scores should be available in the participants’ profiles along with their expertise level in usability testing, which should change as they perform more tests. The amount of guidance should be faded with the increase of expertise.

c. Parents should be provided with some examples of usability problems before the session starts.

d. Instructions should be provided in different modalities in order to cover all preferences. Also, if only video instructions are provided, participants with slow internet connection may not be able to watch or focus while watching the instructions.

e. Video instructions should carefully follow the principles of cognitive load theory. Long videos are not preferable; it is recommended to divide the video logically into shorter videos and to place them right before the related task when possible. Users should also be able to change the quality of the video in case the internet connection is slow.

f. Special attention should be paid to the instructions that help the parent understand who should do what and when. Segmenting the instructions into steps and duplicating them briefly on the child’s device can be helpful.
g. Parents should know exactly what is coming next, and what is their role and the children’s role in each part in order to be prepared before starting.

h. Watching a short example of moderating a usability testing session is recommended for the parents. The parents should be able to type their observations while watching the example to understand and practice their role, and then compare their report with an exemplary report that shows the usability problems discovered in the video.

i. Only necessary instructions should be included. Highlighting and cueing the important parts of the instructions are recommended in both video and written instructions.

j. Parents should be provided with an explanation or a short video that shows the goal of the product to be tested, and how it can be used in order to concentrate on finding problems during the session rather than learning how to use it.

k. If a timer is necessary, an automated timer should be provided for the parent. Also, detailed directions should be provided on when to start the timer, what should the parent do during that time, and what should happen if the child wants to stop, have a break, or change the activity.

l. Parents are recommended to discuss the children’s answers before typing them or letting them type them. They should also be instructed to encourage the children to provide more details.
m. If videotaping is necessary, the parent should be instructed not to place the child behind a light source so that it does not affect the quality of the recording.

n. The parents should be reminded to speak loudly using the language of the report right before recording the session. The children are expected to imitate their parents.

o. Parents should be instructed to help the children provide their honest feedback when discrepancies are noticed between what they really feel and what they say.

p. The parent’s report should include both open ended and more specific tasks and questions in order to give a chance for exploring and discovering new problems, and at the same time to highlight the important parts and guide the parent and the child through the test.

q. Special attention should be paid to the way the questions and tasks are written, which may affect the parents’ and the children’s feedback. For example, placing the word “notes” inside a textbox might indicate that the parent should enter brief notes rather than a detailed observation.

r. Each remote usability testing tool is different in its details; it is important to conduct a pilot test before actually using the remote usability testing tool in order to avoid any problems that could affect the parent’s feedback.

3. Recommendations that are related to the parents’ concerns

a. In order to decide whether the product is appropriate for their children in terms of age and culture, parents should be able to visit the product that
will be tested by their children before deciding whether they want to participate.

b. Videotaping the child’s face is not recommended. If necessary, it is recommended to leave it optional. In this case, the child must know that he or she will be videotaped.

c. It should be stated for the children that the monetary rewards are under the parents’ supervision. Other types of rewards that can be acceptable are free subscriptions in educational products or games, and scores that can be used to build a character, unlock levels, or buy things in the usability testing website itself.

d. Micro-payments are not recommended to be the only payment type in the tool because collecting micro-payments could be addictive for the children. Parents should have different reward options for them and for their children.

e. It is recommended that the remote usability testing tool be authorized by a trusted party to assure the parents that it is safe for children.

Recommendations for Future Research

1. Future research should investigate how different parents’ attributes, such as language, typing skills, usability testing experience, and knowledge about the field of the product to be tested, influence the quality of their feedback in a remote usability testing context.

2. Additional research should examine the children’s preferences and concerns in crowdsourced usability testing platforms.
3. Additional research should investigate the instructional designers’ needs and concerns regarding crowdsourced usability testing with children.

4. Future research should use the guidelines proposed in this study in the design of another instance of a crowdsourced usability testing tool in order to use it with children of different ages using products of different fields.

5. Since this study investigated the use of crowdsourced usability testing of children’s products with the help of the parent as the test moderator, additional research should compare between the usability testing reports provided through tests that are performed by children and their parents with those performed by children, which might be helpful in cases where the parent has limited knowledge with technology or with the field of the product being tested.

Conclusion

Instructional designers of children’s interactive products have not been able to take advantage of the potentials of remote and crowdsourced usability testing. Therefore, it is recommended to develop a remote usability testing tool that is specially designed for testing children’s products using children participants. The current study provided a set of guidelines that can help practitioners and researchers in the design of such platforms. Future research should focus on testing and improving these guidelines and investigating the feasibility of remote and crowdsourced usability testing with children.
References


Klein, J. D. (2014). Design and Development Research: A Rose by Another Name. AERA Design-Based Research Conference, Philadelphia, PA.


Resnik, L. (2011). Development and testing of new upper-limb prosthetic devices: 
Research designs for usability testing. *Journal of Rehabilitation Research &
Development, 48*(6), 697-706.

instructional design and development. *Handbook of research for educational
communications and technology, 2*, 1099-1130. Mahawah, NJ: Laurence Erlbaum 
Associates, Inc.

Richey, R., & Klein, J. D. (2005). Developmental research methods: Creating knowledge 
from instructional design and development practice. *Journal of Computing in

Richey, R., & Klein, J. D. (2014a). Design and development research. In J. M. Spector, 
M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on
Educational Communications and Technology* (pp. 141-150) [Kindle version]. 
New York, NY: Springer.

Richey, R. C., & Klein, J. D. (2014b). Design and development research: Methods,
strategies, and issues [Kindle version]. Routledge.


How to plan, design, and conduct effective tests*. Indianapolis, Ind: Wiley 
Publishing.


Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations* [Kindle version]. Doubleday.


Appendix A

Storyboard

Homepage

---

**Logo**

**CROWDSOURCED USABILITY TESTING OF KIDS’ PRODUCTS**

**Designers/Product developers** Get kids’ feedback and develop better online kids’ products from home.

**Parents** Help product developers create better online educational websites and games for kids all over the world. Try our stay-at-home product-testing.

**Here’s how it works**

1. Product developers post their products to be tested by kids from all over the world.

2. Parents select a product from the list and observe their kids using it.

3. Parents submit their observation notes which are sent to the product developer to enhance his/her product.

---

**Get in touch**

Email: ...........@ohio.edu
Tel: +1..............

---

**Send a message**

Submit
FAQ page

How much time will my child spend on KidCrowd?

How do I guarantee the privacy of my data?

Are the products safe for children?

What are the rewards I can earn?

Get in touch

Email: ..........@chio.edu
Tel: +1............

Send a message

Submit
What is Kid Crowd?

Kid Crowd was created as part of a university instructional technology project with the aim of developing a new technique of testing children's interactive products at home. The project is the beginning of a series of research projects that investigate the potential of crowdsourcing in usability testing of children's educational websites, apps, and games. The ultimate goal is to enhance the overall quality of the products created for children by facilitating the testing procedure, which is critical but often overlooked since it is costly and cumbersome.
Create your profile

Name

e.g. Jamie Smith

Gender

- Male
- Female

Age

Must be 18 years or older

Country of citizenship

How long have you been in the US

How good is your English

Select one

Basic: you can say and understand a few things in English.
Intermediate: you can use English in basic everyday contexts.
Advanced: you speak and understand English with the occasional mistake.
Fluent: you use English with ease and fluency.
Native: English is your first language.

Next
Create your profile

Age of your participating child
- e.g. 9

Gender
- Male
- Female

Have you or your child had any experience with usability testing?
- Explain

How long has your child been using computers?
- e.g. 6 months

Check all the mobile devices your child has been using:
- Laptops
- Tablets
- Smartphones

How good is your child's English

Select one

- Basic: he/she can say and understand a few things in English.
- Intermediate: he/she can use English in basic everyday contexts.
- Advanced: he/she speak and understand English with the occasional mistake.
- Fluent: he/she use English with ease and fluency.
- Native: English is his/her first language.

Save
Parent’s interface

Logo

**Current products**
Select a product to test

<table>
<thead>
<tr>
<th>Categories</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Websites</td>
<td>(2)</td>
</tr>
<tr>
<td>Games</td>
<td>(0)</td>
</tr>
<tr>
<td>Applications</td>
<td>(1)</td>
</tr>
</tbody>
</table>

Title of the first product-testing project
Summary paragraph about the details of the project […]

View More

Title of the second product-testing project
Summary paragraph about the details of the project […]

View More

Get in touch
Email: ……………@ohio.edu
Tel: +1………………

Send a message

Submit
Details of the selected testing project

What is Tynker?
Tynker is an educational platform that teaches coding to children seven years old and above through online games and courses. It meets certain standards of online safety and privacy for children as it has been reviewed and certified by kidSAFE® Seal Program.

What's the task?
Observe your child while competing some tasks on Tynker and tell us how easy it was to use it and what problems were encountered.

Who can participate?
Kids seven or eight years old who have at least six months of experience with computers.

What's in it for me and my child?
Besides being introduced to coding, which is a necessary life skill that teaches creativity and critical thinking, you and your child will earn $10 each to thank you for your time.
The set-up and checklist page

Please read the following instructions carefully:

- Make sure the room is quiet and free from distractions.
- Siblings shouldn't be allowed in the room.
- Use two computers, one for you and one for your child.
- Sit next to your child in such a way that allows you to see his/her screen and face.

Now complete the following check list. In case it is not possible to follow any of the instructions above, please let us know and explain the situation.

- The room is quiet and free from distractions.
- There are no siblings in the room.
- You and your child are using two computers.
- Your seat is next to your child's seat, and you can see his/her screen and face.

Explain

Get in touch

Email: ...........@ohio.edu
Tel: +1....................

Send a message

Submit
The parent’s tour and instructions page

Tour and instructions

Take the tour and see how you can use the website during the session to provide a high-quality report.

When you're ready, click to proceed to the setup page

Get in touch
Email: ........@ohio.edu
Tel: +1................

Send a message

Submit
Starting the usability test

Please open www.kid-crowd.com/kids on your child’s computer and enter the following code to grant him/her access to the website: KC123

If your child accepts to participate in the study, he/she will be performing a warm up activity, then he/she will be directed to the website that needs to be checked.

When your child is directed to Tynker.com, please start recording the session as follows:

[Diagram showing three squares connected by arrows]

Get in touch
Email: .........@ohio.edu
Tel: +1..............

Send a message

Submit
Please read (🌟) the following instructions for your child and fill out (✍️) the report:

1. "Take a look at the site and tell me what you think it is for. You can scroll if you want, but please don't click on anything yet. Just look around and speak everything that comes to your mind."

   Type your child's answer.

2. "View all the activities that are made for kids your age."

   How long did it take your child to find the grades button:
   - ☐ 10 seconds or less
   - ☐ 10 to 60 seconds
   - ☐ He/she didn't find the grades buttons

   Notes

Always give positive feedback even if your child could not complete the task successfully.
3. "Select an activity you'd be interested in doing. Make sure it matches your coding skills."

Did your child notice the beginner/intermediate/advanced categories under the title of each activity?

- Yes
- No
- Yes, but he/she didn't pick an activity that matches his/her coding level

Remember, when your child asks you what to do, redirect the question to him/her and try not to give clues unless ...

4. "You have 10 minutes to play in the activity you picked."

Your child seems like he/she is:

- Having fun
- Frustrated
- Bored
- Engaged
- Distressed

If your child is distressed, please end the testing session.
The quality of the usability test depends on the report you submit. We know how hard it is to type everything you notice in the middle of the observation. Therefore, we are giving you the opportunity to edit your entries before submitting the final report. Please review your report and then click on submit.
Tell us your opinion

Please ask your child to answer the following questions when he/she is done with the activity.

How would you rate Tynker?

Awful  Not very good  Good  Really good  Brilliant

How much did Tynker make coding fun?

Not at all  Just barely  Somewhat  Very much  As much as possible

How easy was it to use Tynker?

Hard  A bit hard  Okay  Quite easy  Really easy

What do you like best about Tynker?

Please type your child's answer here

What do you like least about Tynker?

Please type your child's answer here
How easy was it to understand the activity you picked?

- Hard
- A bit hard
- Okay
- Quite easy
- Really easy

Would you visit Tynker again?

- Yes
- No
- Maybe

Would you play the activity you picked again?

- Yes
- No
- Maybe

Do you want to add anything that can help us make Tynker better?

Please type your child's answer here

Submit
The parent’s gratitude page

We appreciate your time and effort.
To thank you for your time, we would like to send you a $10 gift card. Please enter your email below.

e.g. john@gmail.com

Send the Gift Card

Get in touch

Email: ..........@ohio.edu
Tel: +1.................

Send a message

Submit
The child’s login page

Welcome to Kid Crowd!
Please ask your parent to enter the login code.

Enter the code here

Submit
Hello!

We are doing a study to learn about how to know if a website or computer game is fun and good for kids. We are asking you to help because we don’t know very much about how to make kids your age help us in making better websites for other kids.

If you agree to be in our study, your parent will ask you to do some tasks on a website. We want you to tell us what you think about the website, so your parent will ask you some questions about whether you think the website is fun or not, and why. We also want to know if you think there is anything we can do to make the website better.

A camera and a microphone will record your face and voice while you are using the website so we can watch the video later to write down what you think about the website. Your screen will also be recorded so we can watch it later to know how to make the website better.

You can ask your parents any questions you have about this study. If you decide at any time not to finish, you can ask us to stop.

The questions we will ask are only about what you think. There are no right or wrong answers because this is not a test.

If you type your name and click on the continue button, it means that you have read this and that you want to be in the study. If you don’t want to be in the study, don’t click on continue. Being in the study is up to you, and no one will be upset if you don’t sign this paper or if you change your mind later.

Name: [Type your first name]  
Continue
Before we begin the session, let's have a quick warm up activity with the mouse/touchpad and pointer!

Now, practice moving the pointer onto the bubbles and clicking on them.
Child’s instructions page

Logo

What should I know before I start the session?

Please watch this video to know what will happen next:

When you’re ready to start the session, remind your parent to start recording the session then click on the button below.

Start
Appendix B

Screenshots of the Prototype After the Iterations

Kid-crowd.com homepage 1/2

**Product designers/developers** can receive kids’ feedback on their online educational products from home through Kid Crowd.

**Parents and kids** can help product developers create better online educational websites and games for kids all over the world using our stay at home product testing platform.

Here's how it works

**Product Designers and Developers**

Product designers and developers post their products on Kid Crowd to be tested by kids from all over the world. The products posted are reviewed by Kid Crowd team to make sure that the content is suitable for kids before they are available in Kid Crowd product list.
Kid-crowd.com homepage 2/2

Kids

Kids try out a product, selected by their parents, by performing specific tasks and answering some questions in order to discover any usability problems in the product. Usability problems are any problems in a product, such as a learning website or game, that affect the product’s ease of use, efficiency, or user’s satisfaction. At the end of the session, kids receive gift cards presented from the product designer in return for their time and help.

Parents

After selecting one of the products from Kid Crowd product list to be tried by their kids, parents moderate the session and observe their kids while they are using the product. They fill a report regarding the usability problems discovered during the session based on their kids’ use and comments, and send it to the product designer who reviews the report and makes the necessary changes on his or her product to provide better learning experiences for kids.

Use the sign up button at the top of the page to join Kid Crowd

Get in touch

E-mail: kidcrowd@gmail.com
Phone: +1 (123) 456 7890
Fax: +1 (123) 456 7891

Send a message

Name
Email Address
Subject
Your text here

Submit

© Amirah Alkhawajah 2017. All Rights Reserved.
Kid Crowdsourcing is a platform that was created to support a research project in instructional technology. The aim was to develop a new method for testing children's interactive products at home. Crowdsourcing is an online participatory activity where people seek assistance from a diverse group of individuals around the world to achieve a specific goal that cannot be achieved otherwise at the same level of cost-effectiveness. The project began a series of research projects that explored the potential of crowdsourcing in usability testing of children's educational websites, apps, and games. The ultimate goal was to enhance the overall quality of the products created for children by facilitating the testing process, which is critical but often overlooked since it is costly and cumbersome.
Kid Crowd Frequently Asked Questions

How much time will my child spend on Kid Crowd?

The time your child will spend on trying out a single project depends on the project itself. Some projects require a single session, which is usually between 30 to 60 minutes, while others might require your child to use the product regularly for a specific number of days. The duration of the usability test is provided in the project’s page in which you can read all the details before accepting to participate in the project.

How do I guarantee the privacy of my data?

Are the products safe for children?

What kind of rewards does Kid Crowd offer?
Kid-crowd.com Sign Up page

Sign Up

Product Developer
Username
Email
Password
Re-enter password
Male
Female
Country
Language

Parent Sign Up

Have an account? Log in here

Get in touch
E-mail: kb4950n@ohio.edu
Phone: +1 (323) 456 7890
Fax: +1 (323) 456 7891

Send a message
Name
Email Address
Subject
Your text here

Submit

© Amirah Alkawi 2017. All Rights Reserved.
Designed by Themouth
Create Your Profile

Name
  name

Gender
  ⬜ Male    ⬜ Female

Age
  Must be 18 years or older

Country of Citizenship
  Country of citizenship

How long have you been in the US?

How good is your English
  ⬜ Basic: You can say and understand a few things in English.
  ⬜ Intermediate: You can use English in basic everyday contexts.
  ⬜ Advanced: You speak and understand English with the occasional mistakes.
  ⬜ Fluent: You use English with ease and fluency.
  ⬜ Native: English is your first language.

Age of your participating child
  e.g. 9

Gender
  ⬜ Male    ⬜ Female

Have you or your child had any experience with usability testing before?
  If yes, explain.

How long has your child been using computers?
  e.g. 6 months
Kid-crowd.com Profile page 2/2

Check all the mobile devices your child has been using:
- Laptop
- Tablet
- Smartphone

How good is your child’s language?
- Basic: Can say and understand a few things in English.
- Intermediate: Can use English in basic everyday contexts.
- Advanced: Speaks and understands English with occasional mistakes.
- Fluent: Uses English with ease and fluency.
- Native: English is their first language.

UPDATE PROFILE

Get in touch

Email: as148809@sis.edu
Phone: +1 (123) 456-7890
Fax: +1 (123) 456-7891

Send a message

Name
Email Address
Subject
Your text here

SUBMIT

© Anwar-Abhawajah 2017. All Rights Reserved.
Kid-crowd.com Current Projects page (parent’s interface)

Tyknker - Coding for Kids
Tyknker is an educational website that teaches coding to children seven years old and above through online courses and games. If your child is 8 or 10 years old, you can help us discover what children think about Tyknker. [...]

View More

Science Lab App
The Science Lab App is in its last stage of development. We are looking for kids between the ages of 12-13 to try out our app to make sure it doesn't have any problems before its release. [...]

View More

Solution Math Tutor
We are looking for users between the ages of 11-13 with dyscalculia (math learning difficulties) to try out our website. If these conditions apply to your child, then you could help us improve Solution Math Tutor so that other children with dyscalculia can benefit from the website. [...]

View More
Kid-crowd.com Project Details page (for Tynker project)

What is Tynker?
Tynker is an educational platform that teaches coding to children seven years old and above through online courses and games. It meets certain standards of online safety and privacy for children as it has been reviewed and certified by KidSAFE® Seal Program.

What is the task?
You will be asked to observe your child while he or she completes some tasks on Tynker. We want you to tell us how easy it was to use Tynker and what problems were encountered. We will provide you with more details and instructions at the beginning of the session.

Who can participate?
Kids nine or ten years old who have at least six months of experience with computers.

What’s in it for me and my child?
Besides being introduced to coding, which is a necessary life skill that teaches creativity and critical thinking, you and your child will earn $10/week to thank you for your time.

Join the Crowd
Select Another Project

Comments

Post by Tynker
Thank you for your interest in our project. Yes, we encourage kids with any level of coding experience to participate in our project.

Post by Tynker
My kid is a beginner coder. Can he participate?

Post by Tynker

Post by Tynker

Post by Tynker

Post by Tynker
Kid-crowd.com Tour and Instructions (1/3) page

Tour and Instructions (1/3)

For written instructions: click here

1- Parent’s Role

- Example of Your Role -

Please watch the example below of a mother and her daughter performing a usability test, and try to type all the problems she has faced while using the website.

Types in your observations:

Click here to reveal the mother’s report, and compare it with yours.

CLICK TO PROCEED TO PART 2 OF THE INSTRUCTIONS
Kid-crowd.com Report Example page

Usability Testing Report - Example

This was part of the mother's report from the video you just watched:

- My daughter did not know what the website is about.
- The 'Learn More' button does not work or maybe the website is too slow.

Please close the window to go back to the previous page.
Tour and Instructions (2/3)

For written instructions: click here

2- How to Use Kid Crowd

CLICK TO PROCEED TO PART 3 OF THE INSTRUCTIONS
Kid-crowd.com Tour and Instructions (3/3) page

Tour and Instructions (3/3)

For written instructions: click here

3-Tips

3

Tips

CLICK TO PROCEED TO THE SETUP PAGE
Set up the Testing Room

Please read the following instructions carefully:

- Make sure the room is quiet and free from distractions.
- Siblings shouldn’t be allowed in the room.
- Use two computers, one for you and one for your child.
- Sit next to your child in such a way that allows you to see his/her screen and face, and do not let your child sit in front of a light source (window, lamp, etc.).

Now complete the following checklist. In case it is not possible to follow any of the instructions above, please let us know and explain the situation.

☐ The room is quiet and free from distractions.
☐ There are no siblings in the room.
☐ You and your child are using two computers.
☐ Your seat is next to your child’s seat where you can see his/her screen and face, and your child is not sitting in front of a light source.

Explain:

[Blank field]

START THE SESSION
Start the Session

Please Follow the Steps Below to Start the Session

<table>
<thead>
<tr>
<th>Step 1: Child's assent and instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ask your child to go to <a href="http://www.kid-crowd.com/kids.htm">www.kid-crowd.com/kids.htm</a> on his/her laptop. Let him/her enter the following code to be able to access the website: <strong>KC123</strong></td>
</tr>
<tr>
<td>Let your child read the assent form. If your child accepts to participate in the study, he or she will perform a warm-up activity and watch some instructions.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step 2: Recording the session</th>
</tr>
</thead>
</table>

| Step 3: Reading the first task |

NEXT | REPORT 1/2 |
Start the Session

Please Follow the Steps Below to Start the Session

*Step 1: Child's access and instructions*

*Step 2: Recording the session*

Note that your child is alone watching the video, please record the session following the steps below before a file is submitted to Thinker.com to record the session.

1. Go to the desktop, your desktop, and click on the icon of Screenshot software which will be used to record the session.

2. The window below will appear. Click on "Kid Crowd Header".

3. The window below will appear. Fill in the details and the session's name, and fill the rest of the details then click on "Save Changes".

4. Make sure that the file and the parent working time click on "Begin Session".

5. Click on "Start Recording". The recording will start after the countdown. Please speak loudly and clearly because we need to hear your voices as well.

*Step 3: Reading the final task*
Please Follow the Steps Below to Start the Session

1. Child's assent and instructions
2. Recording the session
3. Reading the first task

Now that you have started recording the session, let your child go back to his/her internet browser, and ask him/her to wait until you read the first task from the next page below. (Click on the button below to read the first task.)
Part 1: Please read each task, observe your child, and type your observations:

- **Read to your child:** "Imagine that your friend told you about a website called 'Tyke' and that you wanted to try it out. Take a look at the website and tell me what you think it is for. You can scroll if you want, but please don't click on anything yet. Just look around and speak everything that comes to your mind."

  - **Type the answer:**

- "How are the activities that are made for kids your age?"

  - **How long did it take your child to find the grades button?**

  - [ ] 10 seconds or less.
  - [ ] 10 to 60 seconds.
  - [ ] More than 60 seconds, or he/she didn't find the grades buttons.

- "Select an activity you'd be interested in doing. Please select any activity other than the 'Game' since it doesn't have coding. Just make sure that the one you pick matches your coding skills."

  - **Did your child notice the beginner/intermediate/advanced categories under the title of each activity?**

  - [ ] Yes.
  - [ ] No.
  - [ ] No, but he/she didn't pick an activity that matches his/her coding level.

- "You can start playing now. Try to think out loud while you're playing."

  - **How your child to play at least until your figure out if there are any problems in the game. Type any problems in the text box below. Your child seems like he/she is:

    - [ ] having fun
    - [ ] frustrated
    - [ ] bored
    - [ ] engaged
    - [ ] distressed

- **Did your child face any technical problems while using Tyke or the activity he/she picked?**

  - e.g. unclickable or confusing buttons, activity is too hard or too easy, unclear directions, slow download, inappropriate choice of color, etc.

  - [ ]

Please review your report, adding more details if any, and then click on submit.
An example of the support provided in the first Report page

- “View all the activities that are made for kids your age.”
- How long did it take your child to find the grades button?

![Image of Hour of Code activity]

- “Select an activity you'd be interested in doing. Please note that this is not the same as Minecraft since it doesn't have coding. Just make sure that the one you pick matches your coding skills.”

- Always give positive feedback even if your child couldn't complete a task successfully. This will actually help us discover the website's flaws!
Part 2: Let Your Child Type His Answers

When your child is done with the activity, please let him/her type his/her own answers to the following questions.

How would you rate Tyker?
- Awesome.
- Not very good.
- Good.
- Really good.
- Awesome.

How much did Tyker make coding fun?
- Not at all.
- Just barely.
- Somewhat.
- Very much.
- As much as possible.

How easy was it to use Tyker?
- Kind.
- A bit hard.
- Okay.
- Quite easy.
- Really easy.

What does your child like best about Tyker?
Type your answer here in detail.

What does your child like least about Tyker?
Type your answer here in detail.

How easy was it to understand the activity you picked?
- Kind.
- A bit hard.
- Okay.
- Quite easy.
- Really easy.

Would you use Tyker again?
- Yes.
- No.
- Maybe.

Would you play the activity you picked again?
- Yes.
- No.
- Maybe.

In your opinion, how can we make Tyker better?
You're the expert here. Give us as much detail as you can so we can make Tyker better.

Please stop the recording on your child’s laptop by clicking on the Stop button on the top of the screen, then click on the button below.
Kid-crowd.com Help page

During the session, if your child is...

- **Not Thinking Out Loud**
  - Encourage your child to verbalize his or her thoughts through a dialogue.

- **Distressed for Any Reason**
  - Stop the session immediately. Your child will still receive a gift card for his or her time and help. Never force your child into continuing the session.

- **Having Difficulties Reading Something**
  - Read for your child but never try to complete the task for him or her unless he or she is really stuck.

- **Forgot the Task**
  - Remind the task for your child if it seems like he or she has forgotten the task.

- **Distracted or Bored**
  - Gently remind your child to complete the task or pretend that you need their help to learn how to use the website or complete the task. You can also give your child a short break if the session was long.

- **Tired**
  - Give your child a short break if he or she doesn’t want to continue the session. Never force him or her. Your child will still receive a gift card.

- **Asking for Help**
  - If your child is asking “How can I do that?” ask him or her back, “What do you think? Try your best to do the task for him or her.”

- **Stuck in a Task**
  - If your child couldn’t figure out a task, that could mean that he or she has just discovered a usability problem. Please add it to the report, and help your child complete the task. Always provide positive feedback that doesn’t indicate that he or she completed or failed to complete the required task.
Welcome to Kid Crowd!

Please ask your parent to enter the login code.

Enter the code here

Submit
Hello!

We are doing a study to learn about how to know if a website or computer game is fun and good for kids. We are asking you to help because we don’t know very much about how to make kids your age help us in making better websites for other kids.

If you agree to be in our study, your parent will ask you to do some tasks on a website. We want you to tell us what you think about the website, so your parent will ask you some questions about whether you think the website is fun or not, and why. We also want to know if you think there is anything we can do to make the website better.

A camera and a microphone will record your face and voice while you are using the website so we can watch the video later to write down what you think about the website. Your screen will also be recorded so we can watch it later to know how to make the website better.

You can ask your parent any questions you have about this study. If you decide at any time not to finish, you can ask us to stop.

The questions we will ask are only about what you think. There are no right or wrong answers because this is not a test.

If you click on the continue button and sign the paper copy of this form, which your parent will give you, it means that you have read this form and that you want to be in the study. If you don’t want to be in the study, don’t sign or click on continue. Being in the study is up to you, and no one will be upset if you don’t want to be in it. You can stop at any time if you want to. Please sign the paper copy that is with your parent and then click on continue if you agree to be in the study.
Kid-crowd.com Warm-up Activity page
Kid-crowd.com Kids’ Instructions page

What should I know before I start the session?

Please watch the video below to know what is going to happen next:

NEXT

© Amrah Akhraswijn 2017. All Rights Reserved.
Kid-crowd.com Begin the Session page

Are you ready?

You will go to Tynker.com to tell us how to make it better. We need your sincere opinion, but first:

1- Remind your parent to start recording the session from your desktop.

2- Wait until your parent reads the first task for you from his/her laptop, then click on the button below and start performing the task.

GO TO TYNKER

© Amira Alshawajah 2017. All Rights Reserved.
Kid-crowd.com Thank You page

Thank you!

We appreciate your time and effort. To thank you for your time, we would like to send a $10 gift card for you and another for your parent. Please ask your parent to enter his or her email below.

Email: [_email]

Send the Gift Card

Get in touch

E-mail: [email]
Phone: +1(123) 456-7890
Fax: +1(123) 456-7891

Send a message

Name
Email Address
Subject
Your text here

Submit

© Amirah Alrawi 2017. All Rights Reserved.
Appendix C

Reply Message from a 4th Grade Teacher

On Sat, Nov 18, 2017 at 12:02 PM, [the teacher] wrote:

Amira,

I had the opportunity to look over the web pages you forwarded to me. Following are my opinions.

First, it was such a thrill to hear [a name of a student] voicing over for the characters in the study! She did a fabulous job!

In addition, I found it very user friendly, appropriate, easy to understand, and easy to follow. The font seemed appropriate, and there were no terms or tasks that seemed beyond a 4th grader's ability to understand.

I had a few issues, though, and forgive me, but I'm not very tech savvy; these issues may seem very elementary. Does the participant HAVE to use a laptop computer? I don't own one, so I used my desktop computer. However, the part about using the "mouse" with or without the buttons was a bit confusing to me. I chose the "mouse" with the buttons, but then later on in the tutorial it asked me to push the one big button, and the graphic showed only the one type of "mouse," the one without the buttons, so again, I was a bit confused about what to do.

Another minor difficulty I encountered was when I was instructed in the tutorial to click the red circle to click out of the screen. The YouTube icon was also on the screen, and it also has a red circle. Students are very good at clicking around a site until something happens, so they would probably figure it out. However, it might be a bit clearer to say
to click on the red circle with the X to click out of the screen instead of just clicking on the red circle.

The last minor difficulty I had was when you provided an example of how to interact with the website. The volume suddenly dropped significantly overall, and it was a bit hard to hear what was being modeled.

All in all I was very impressed! I really liked the way you always clearly introduced what was coming up so there were no surprises. I also really appreciated what you said in the very beginning about no one being upset if a student/parent decided to discontinue. All of it had a very gentle, inviting feel, and I know students would enjoy it!

Blessings!

On Mon, Nov 13, 2017 at 4:49 PM, Alkhawajah, Amira wrote:

Sure! I know how busy you are and I really appreciate your help and the time you will spend to complete the task, [the teacher’s name]. Thank you so much!

Kind regards,

Amira Alkhawajah

On Mon, Nov 13, 2017 at 4:22 PM, [the teacher] wrote:

Thank you for your email. Yes, I will be happy to help you. However, I can't commit to 30-60 minutes to complete the task until this weekend. We have parent/teacher conferences this week, and all my outside of school time is taken up with that at the moment. Can you wait till the weekend?
Dear Mrs. [teacher’s name],

I am a student in the Instructional Technology PhD Program at OU, and I would like to ask you for your help in voluntarily evaluating a few web pages I have made for my dissertation titled "Guidelines for Remote Usability Testing of Children’s Interactive Products". I will be working with many 4th and 5th grade students in this study, but before I get started, I need to make sure that the website the children will be using is suitable for their age. I believe that your experience as a 4th grade teacher can help me a lot in discovering whether the word choice, the font and font size, and the media used in the website are suitable for children 9 to 10 years old. Therefore, I would like to ask you for your help. Below is some information to help you make an informed decision:

Did the researcher get permission to conduct this study?

Yes, I have a formal letter from ACSD’s superintendent to use the help of elementary school teachers in my research, and I have obtained OU’s IRB permission.

How many web pages will I evaluate?

You will evaluate 3 web pages.

1. The 1st has the assent form.
2. The 2nd has a tutorial for using the laptop’s touchpad.
3. The 3rd has a 5-minutes video explaining the study to the students.

How much time is it going to take?

I am expecting it to take around 30 to 60 minutes.
What will I do?

If you accept to help in this research, you will do the following:

1. Kindly go to www.Kid-Crowd.com/kids.html and enter the code KC123 to enter the website.

2. In the next 3 web pages, read the assent form, use the touchpad tutorial, and watch the video.

3. Finally, send me an email (aa148809@ohio.edu) to let me know if there are any words or sentences that might be difficult for children 9 to 10 years old to understand. Also let me know if you think the font and font size used in the website is good for this age range. If you notice any other problems that might make the website and/or media hard for the children to use or understand, please let me know.

If you have any questions, please do not hesitate to contact me. I am looking forward to hearing from you.

Kind regards,

Amirah Alkhawajah
Appendix D

Reply Message from Dr. Paas

On Tues, December 5, 2017 at 1:33 AM, Paas, Fred wrote:

Dear Ami rah,

I had a look at the videos and I can only say that they look good. I don’t see any violations against cognitive load theory. I might be able to say more about it if you could tell me how you have used the theory, but anyway, even then I am sure that it just looks good and age-appropriate for the children and parents.

Best wishes,

Fred

Fred Paas, Ph.D. | Professor of Educational Psychology
Department of Psychology, Education & Child Studies | Erasmus University Rotterdam | Netherlands | paas@fsw.eur.nl
Early Start | University of Wollongong | Australia | fredp@uow.edu.au

On Mon, December 4, 2017 at 11:52 PM, Alkhawajah, Amirah wrote:

Dear Prof. Paas,

I was honored to read your articles about cognitive load and instructional design, which I found to be very helpful in creating the product I designed for my dissertation. I am a PhD candidate in instructional technology at Ohio University, and I am working on my dissertation that investigates the use of remote/crowdsourced usability testing of children's interactive learning products.
I used the principles of cognitive load theory in creating the multimedia that are included in the usability testing platform I designed, and I would truly appreciate your help in reviewing them to make sure that they indeed follow the cognitive load theory principles. The items that need to be reviewed are one interactive tutorial that takes around 3 to 4 minutes to complete, and three videos, one of which is around 1 minute long, and the other two videos are around 5 minutes long each.

Here are the links for the items that need to be reviewed in case you kindly decided to help me with your experience and knowledge:

- A one-minute video: http://kid-crowd.com

I am looking forward to hearing from you, and I appreciate the time you took to read my email. If for any reason you won't be able to assist me, please let me know.

Kind regards,

Amirah Alkhawajah
Appendix E
IRB Approval Letter

<table>
<thead>
<tr>
<th>Project Number</th>
<th>17-X-212</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Status</td>
<td>APPROVED</td>
</tr>
<tr>
<td>Committee:</td>
<td>Social/Behavioral IRB</td>
</tr>
<tr>
<td>Compliance Contact:</td>
<td>Robin Stack (<a href="mailto:stack@ohio.edu">stack@ohio.edu</a>)</td>
</tr>
<tr>
<td>Primary Investigator:</td>
<td>Amirah Alkhawajah</td>
</tr>
<tr>
<td>Project Title:</td>
<td>Guidelines for Remote Usability Testing of Childrens Interactive Products</td>
</tr>
<tr>
<td>Level of Review:</td>
<td>EXPEDITED</td>
</tr>
</tbody>
</table>

The Social/Behavioral IRB reviewed and approved your amendment of the above referenced research.

<table>
<thead>
<tr>
<th>IRB Approved:</th>
<th>10/11/2017 4:38:51 PM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expiration:</td>
<td>08/09/2018</td>
</tr>
<tr>
<td>Review Category:</td>
<td>6,7</td>
</tr>
</tbody>
</table>
Waivers: No waivers are granted with this approval.

If applicable, informed consent (and HIPAA research authorization) must be obtained from subjects or their legally authorized representatives and documented prior to research involvement. In addition, FERPA, PPRA, and other authorizations must be obtained, if needed. The IRB-approved consent form and process must be used. Any changes in the research (e.g., recruitment procedures, advertisements, enrollment numbers, etc.) or informed consent process must be approved by the IRB before they are implemented (except where necessary to eliminate apparent immediate hazards to subjects).

The approval will no longer be in effect on the date listed above as the IRB expiration date. A Periodic Review application must be approved within this interval to avoid expiration of the IRB approval and cessation of all research activities. All records relating to the research (including signed consent forms) must be retained and available for audit for at least three (3) years after the research has ended.

It is the responsibility of all investigators and research staff to promptly report to the Office of Research Compliance / IRB any serious, unexpected and related adverse and potential unanticipated problems involving risks to subjects or others.

This approval is issued under the Ohio University OHRP Federalwide Assurance #00000095. Please feel free to contact the Office of Research Compliance staff contact listed above with any questions or concerns.
IRB Approval Letter After Amendment

<table>
<thead>
<tr>
<th>Project Number</th>
<th>17-X-212</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Status</td>
<td>APPROVED</td>
</tr>
<tr>
<td>Committee:</td>
<td>Social/Behavioral IRB</td>
</tr>
<tr>
<td>Compliance Contact:</td>
<td>Robin Stack (<a href="mailto:stack@ohio.edu">stack@ohio.edu</a>)</td>
</tr>
<tr>
<td>Primary Investigator:</td>
<td>Amirah Alkhawajah</td>
</tr>
<tr>
<td>Project Title:</td>
<td>Guidelines for Remote Usability Testing of Childrens Interactive Products</td>
</tr>
<tr>
<td>Level of Review:</td>
<td>EXPEDITED</td>
</tr>
</tbody>
</table>

The Social/Behavioral IRB reviewed and approved by expedited review the above referenced research. The Board was able to provide expedited approval under 45 CFR 46.110(b)(1) because the research meets the applicability criteria and one or more categories of research eligible for expedited review, as indicated below.

<table>
<thead>
<tr>
<th>IRB Approved:</th>
<th>01/26/2018 10:30:15 AM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expiration:</td>
<td>08/09/2018</td>
</tr>
<tr>
<td>Review Category:</td>
<td>6,7</td>
</tr>
</tbody>
</table>
Waivers: No waivers.

If applicable, informed consent (and HIPAA research authorization) must be obtained from subjects or their legally authorized representatives and documented prior to research involvement. In addition, FERPA, PPRA, and other authorizations / agreements must be obtained, if needed. The IRB-approved consent form and process must be used. Any changes in the research (e.g., recruitment procedures, advertisements, enrollment numbers, etc.) or informed consent process must be approved by the IRB before they are implemented (except where necessary to eliminate apparent immediate hazards to subjects).

The approval will no longer be in effect on the date listed above as the IRB expiration date. A Periodic Review application must be approved within this interval to avoid expiration of the IRB approval and cessation of all research activities. All records relating to the research (including signed consent forms) must be retained and available for audit for at least three (3) years after the research has ended.

It is the responsibility of all investigators and research staff to promptly report to the Office of Research Compliance / IRB any serious, unexpected and related adverse and potential unanticipated problems involving risks to subjects or others.

This approval is issued under the Ohio University OHRP Federalwide Assurance #00000095. Please feel free to contact the Office of Research Compliance staff contact listed above with any questions or concerns.
Appendix F

Observation Guide

Date:

Time:

Setting description (room, people, devices)


1- Set-up and planning

Start time: _____ am/pm

Participants’ actions:


Participants’ attitude:


Notes:


End time: _____ am/pm   Duration: _____ minutes
2- Introduction

Start time: _____ am/pm

Participants’ actions:

________________________________________________________________________

________________________________________________________________________

Participants’ attitude:

________________________________________________________________________

________________________________________________________________________

Notes:

________________________________________________________________________

________________________________________________________________________

End time: _____ am/pm   Duration: _____ minutes

3- During the test

Start time: _____ am/pm

Participants’ actions:

________________________________________________________________________

________________________________________________________________________

Participants’ attitude:

________________________________________________________________________
4- Finishing up

Start time: _____ am/pm

Participants’ actions:

Participants’ attitude:

Notes:

End time: _____ am/pm   Duration: _____ minutes
Hello! My name is Amirah Alkhawajah. I’m an instructional technology student from Ohio University. As you might know, a lot of children’s websites and applications are not properly tested before being released, which results in defective products. In this study, we propose crowdsourcing as a new approach to testing children’s interactive products in order to help the designer test the product more efficiently before releasing it to the public. Crowdsourcing means using the help of a group of people to solve a problem or complete a specific task through the internet. The purpose of this interview is to discover how to improve the parent’s role in crowdsourced usability testing and to understand your opinion and your concerns regarding your child’s participation in such activities. There are no right and wrong answers. I would like you to feel comfortable saying what you really think and how you really feel. If it’s okay with you, I will be audio recording our conversation in order to transcribe it later. Everything you say will remain confidential.

I believe you have read, understood, and signed the informed consent, right?

**Background data**

Could you tell me about yourself?

- Education level.
- Occupational status.
- Computer use.
• Number of children.

**Parent’s role (reducing workload and maintaining good quality)**

• How would you describe the experience you had today with the usability testing session?

• To what degree do you feel you have accomplished the goal of the session?
  - Probe: What do you think about the quality of the feedback you provided during the session?
  - To what extent did your experience with technology and understanding of the games your child played affect your feedback and your ability to find problems and describe them?
  - How did your English language affect your understanding of the instructions and the quality of the feedback you provided? (this question is only for non-English native speakers)
    - Probe: (If bad): Why?

• What is your opinion of the instructions section?
  - Probe: What would you change?
  - Probe: Do you think the instructions provided prior to the test was enough to carry a successful session?

• What is your opinion of the support provided in the parent’s page?
  - Probe: Did you use it?
  - Probe: What would you change?
  - Probe: How helpful was the practice section before starting the real session?
• What kinds of problems have you faced during the session?
  o Probe: Why do you think this happened?
  o Probe: How about:
    ▪ Understanding your role in the usability test?
    ▪ Determining usability problems?
    ▪ Workload?
    ▪ Dealing with the child’s reactions?
• In the usability testing session that you have experienced today, you took the role of the test administrator (preparing the setting and the equipment and explaining the test procedures), the technical troubleshooter (solving technical problems), the observer (watching, listening, and taking notes), and the interviewer (asking the child some questions and typing the answers). Could you tell me which roles were harder to you?
  o Probe: Why do you think this specific role was hard?
• What improvements do you recommend for the usability testing website?
• How important is the parent's presence in online usability testing sessions?
  o Probe: Do you think children this age can perform an online usability testing session by their own?

**Children-related features**
• Some features have been suggested from the last testing session and were implemented in the new version of the website. What do you think of each feature? Examples of potential features/characteristics:
  o Giving the child a break in the middle of the session.
Typing the answers for the child vs allowing the child to type the answers.

The timing of asking the child some questions about the usability of the tested product.

- At a specific point in the session your child seemed bored / distracted / not interested. Can you tell me why?
  - Probe: How do you think the website can be improved to avoid this from happening?

**Parent’s concerns (safety, time, reward)**

- What kinds of concerns do you think parents might have regarding their children’s participation in online usability testing?

- What could be some negative or positive effects on the child being a participant in online usability testing sessions?

- Safety and privacy.
  - What are your concerns regarding safety and privacy?
  - What types of recordings would be acceptable/unacceptable (screen, audio, video). Why?

- Time and frequency.
  - In your opinion, how long should a usability testing session be for a comfortable experience for a parent and his or her child?
  - How frequently would you let your child use such platforms?
    - Probe: (If not much) What if the monetary reward was replaced with a different kind of reward?

- Reward.
- What types of rewards do you expect your child to earn in such platforms?
- What are your concerns regarding monetary rewards?
- What is your opinion of providing mini-payments that are paid for mini-tasks in place of the typical payment paid for a whole testing session?
Appendix H

Consent Form/Parental Consent Form

Title of Research: Testing Children’s Interactive Products

Researchers: Amirah Alkhwajah

You are being asked permission for your child to participate with you in research. For you to be able to decide whether you want to participate in this project, you should understand what the project is about, as well as the possible risks and benefits in order to make an informed decision. This process is known as informed consent. This form describes the purpose, procedures, possible benefits, and risks. It also explains how your and your child’s personal information will be used and protected. Once you have read this form and your questions about the study are answered, you will be asked to sign it. By signing this form, you agree to participate, and you give us permission to ask your child to participate in the study. You should receive a copy of this document to take with you.

Explanation of Study

This study is being done to develop practical guidelines for testing the quality of children’s educational website. If you agree to participate and allow your child to participate, your child will be asked to use a website that teaches coding in order to know how good the website is and how to improve it. The session will be done at your home. The researcher will visit your home to observe the session and take notes. An audio, webcam, and screen recorders will be used during the session.
This means that your voices, faces, and your screens will be recorded in order to take notes later after the end of the session.

Before your child starts using the website, he or she will perform an online warm-up activity to practice the input devices (e.g. computer mouse, touchpad) before the beginning of the test. Your child will then receive some online instructions on how to help us check the quality of the website. You will also be given some instructions on how to monitor the session and how to report the results through a website especially designed for this study. On that website, you will read some tasks for your child to follow. You will observe your child performing the tasks on the coding website to discover any problems in its design. After each task, you will be provided with some questions that your child will answer to help us discover any problems in the design of the coding website. At the end of the session, you will be asked to review the answers and the comments you provided and submit the report. The researcher will then interview you to gain a deeper understanding on how to improve the quality assurance process.

Your child should not participate in this study if he or she:

- is not nine or ten years old;
- does not have at least six months of experience with computers;

Your child’s participation in the study will last for about one hour, and your participation will last for about two hours.
Risks and Discomforts

No risks or discomforts are anticipated.

Benefits

This study is important to society because it will help in providing better educational websites and applications for children by providing the designers of such products with a practical way of assuring the quality of their products. Individually, your child may benefit by being introduced to a website that teaches coding, which is a skill that teaches creativity, logical thinking, and problem solving.

Confidentiality and Records

Your study information will be kept confidential by:

- Storing study data and documents in password protected files;
- Replacing direct identifiable data with codes;
- Making the recordings available for the research team only;
- Destroying the recordings in August 2020.

Additionally, while every effort will be made to keep your child’s study-related information confidential, there may be circumstances where this information must be shared with:

- Federal agencies, for example the Office of Human Research Protections, whose responsibility is to protect human subjects in research;
• Representatives of Ohio University (OU), including the Institutional Review Board, a committee that oversees the research at OU;

Compensation

As compensation for your time and effort, you and your child will receive $10 gift cards.

Please be aware that certain personal information, such as name, address and social security number, may be provided to the Ohio University Finance Office to document that your child received payment for research participation. However, your child’s study data will not be shared with Finance.

Contact Information

If you have any questions regarding this study, please contact the investigator Amirah Alkhawajah, aa148809@ohio.edu, (740)707-0516, or the advisor Dr. David Moore, moored3@ohio.edu.

If you have any questions regarding your child’s rights as a research participant, please contact Dr. Chris Hayhow, Director of Research Compliance, Ohio University, (740)593-0664 or hayhow@ohio.edu.

By signing below, you are agreeing that:
• you have read this consent form (or it has been read to you) and have been given
the opportunity to ask questions and have them answered;

• you have been informed of potential risks to your child and they have been
explained to your satisfaction;

• you understand Ohio University has no funds set aside for any injuries your child
might receive as a result of participating in this study;

• you are 18 years of age or older;

• your child’s participation in this research is completely voluntary;

• your child may leave the study at any time; if your child decides to stop
participating in the study, there will be no penalty to your child and he/she will
not lose any benefits to which he/she is otherwise entitled.

Parent Signature_________________________________________ Date___________

Printed Name______________________________________________

Child’s Name_______________________________________________

Version Date: 09/11/17
Appendix I

Assent Form*

We are doing a study to learn about how to know if a website or computer game is fun and good for kids. We are asking you to help because we don’t know very much about how to make kids your age help us in making better websites for other kids.

If you agree to be in our study, your parent will ask you to do some tasks on a website. We want you to tell us what you think about the website, so your parent will ask you some questions about whether you think the website is fun or not, and why. We also want to know if you think there is anything we can do to make the website better.

A camera and a microphone will record your face and voice while you are using the website so we can watch the video later to write down what you think about the website. Your screen will also be recorded so we can watch it later to know how to make the website better.

You can ask your parent any questions you have about this study. If you decide at any time not to finish, you can ask us to stop.

The questions we will ask are only about what you think. There are no right or wrong answers because this is not a test.

If you click on the continue button and sign the paper copy of this form, which your parent will give you, it means that you have read this form and that you want to be in the study. If you don’t want to be in the study, don’t sign or click on continue. Being in the study is up to you, and no one will be upset if you don’t want to be in it. You can stop at any time if you want to. Please sign the paper copy that is with your parent and then click on continue if you agree to be in the study.
Your name: ______________________________ Signature:__________

*Adapted from (https://www.irb.cornell.edu/forms/assent.htm).

Version Date: **07/31/17**
Appendix J

Iterations’ Details

First Iteration

The first iteration included the following:

- Five-minute long instructions videos. One for the parent and one for the child.
- The child’s video included:
  - the goal of the session;
  - what is going to happen next;
  - and an example of a usability testing session of a product that is similar to the one in the current study and with similar tasks.
- The parent’s video included:
  - the parent’s role in the session;
  - how to use Kid Crowd during the session;
  - and some tips that were derived from Hanna’s guidelines that might be helpful during the session.
- Parents were asked to type both the first part and the second part of the report.
- In one of the tasks, the children were told that they should play for at least ten minutes in an activity of their choice. No timer was provided; the parents were supposed to keep track of the time.

Second Iteration

The second iteration included the following:
• The set-up instructions included one more point, which is making sure that the child does not sit behind a light source so it does not affect the quality of the recording.
• Written instructions were added; parents had the choice to watch or read the instructions.
• The parents’ instructions video was five minutes long. It included:
  o the parent’s role in the session;
  o how to use Kid Crowd during the session;
  o and some tips that were derived from Hanna’s guidelines that might be helpful during the session.
• The example of a usability testing session was removed from the child’s video. The video was almost two minutes long, and it included:
  o the goal of the session
  o and what is going to happen next.
• The parent’s instructions on how to record and start the session were divided into steps, and reminders of these steps were briefly duplicated in the child’s screen.
• Cueing and visual reminders were used in the text, screenshots, and videos.
• The parents were told exactly what is expected from them in the report pages right before they opened them. The titles and the buttons of the report pages indicated how many pages there are and what exactly each report page is about: Observation or interview.
• Parents were asked to type both the first part and the second part of the report.
• In the last task, the children were told that they have 15 minutes to play in an activity of their choice. A timer was provided for the parents to keep track of the time.

• The wording of the tasks/questions in the report pages were changed. For example, the word “notes” was replaced with a sentence that asks the parent to type his or her observations in details with some examples of usability testing problems. Children’s questions were changed from “do you want to add anything that can help us make this product better?” to “in your opinion, how can we make this product better?”

• A webpage that only has a text note was placed between the usability testing prototype and the product to be tested in order to make the child realize that they are not the same website and that the testing session starts after that page.

Third Iteration

The third iteration included the following:

• Parents had the choice to watch or read the instructions.

• The parents’ instructions video was divided into three short videos:
  o A video that explains the parent’s role in the session.
  o A video that explains how to use Kid Crowd during the session.
  o A video that provides some tips that were derived from Hanna’s et al. (1997) guidelines that might be helpful during the session.

• The first video that talks about the role of the parent was accompanied by another video that was almost one minute long, and it provided an example of a parent
moderating a usability testing session. The parent was supposed to watch the video and practice typing the report at the same time.

- A new example of a usability testing session was added to the child’s video. The video was almost three minutes long, and it included:
  - the goal of the session;
  - what is going to happen next;
  - and an example of a usability testing session of a product that is totally different than the product to be tested by the child in the current study, and it demonstrated negative feedback as well as positive feedback.

- The parent’s instructions on how to record and start the session were divided into steps, and reminders of these steps were briefly duplicated in the child’s screen.

- Cueing and visual reminders were used in the text, screenshots, and videos.

- The parents were told exactly what is expected from them in the report pages right before they opened them. The titles and the buttons of the report pages indicated how many pages there are and what exactly each report page is about: Observation or interview.

- Parents were asked to type the first part of the report only in which they observe and type their observations. The children were asked to type their answers to the second part of the report, which asks them about their opinion in the product.

- There were no time restrictions or instructions.

- The text boxes that followed the multiple-choice questions in the observation report page were removed. The only text boxes that were in the observation report page were the ones that followed the open-ended questions/tasks.