Challenges of Replicating Empirical Studies with Children in HCI

Quincy Brown
Games+Mobile Play Learn Live Lab
Bowie State University
14000 Jericho Park Road
Computer Science Building
Bowie, MD 20715 USA
qbrown@bowiestate.edu

Lisa Anthony
UMBC
Information Systems
1000 Hilltop Circle
Baltimore, MD 21250 USA
lanthony@umbc.edu

Robin Brewer
UMBC
Information Systems
1000 Hilltop Circle
Baltimore, MD 21250 USA
brewer3@umbc.edu

Germaine Irwin
UMBC
Information Systems
1000 Hilltop Circle
Baltimore, MD 21250 USA
germaine.irwin@umbc.edu

Jaye Nias
Games+Mobile Play Learn Live Lab
Bowie State University
14000 Jericho Park Road
Computer Science Building
Bowie, MD 20715 USA
jayeaclark@aol.com

Berthel Tate
Games+Mobile Play Learn Live Lab
Bowie State University
14000 Jericho Park Road
Computer Science Building
Bowie, MD 20715 USA
TATEB0528@students.bowiestate.edu

Abstract
In this paper, we discuss the challenges of conducting a direct replication of a series of mobile device usability studies that were originally conducted with adults and older children (ages 7 to 17). The original studies were designed to investigate differences in how adults and children use mobile devices to touch targets and create surface gestures. In this paper, we report on a replication we conducted with young children (ages 5 to 7). We discuss several methodological changes that were needed to elicit the same quality of data from the replication with young children as had been obtained from the older children and adults. The insights we present are relevant to the extension of empirical studies in HCI in general to younger children.

Author Keywords
Child-computer interaction, touch interaction, gesture interaction, mobile devices, replication, empirical study.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Design, Human Factors.

Introduction

In the context of research studies, children have often been viewed as vulnerable or ill-equipped and have been excluded from participation in studies due to concerns regarding informed consent, confidentiality, and the specialized attention or procedures required when conducting research with minors [1, 2]. To that end, previous research in human-computer interaction (HCI) has focused on including child participants by developing child-centered research methods and adapting protocols specifically for children [1, 3]. Though these efforts have increased the inclusion of children in HCI research, the use of separate protocols does not always allow for the direct comparison of findings between adults and children.

Here we present our insights from a direct replication of a series of studies of touch and gesture interaction on mobile devices that was first conducted with adults (18 years and older) and older children (ages 7 to 17) [6, 7, 8], and then replicated with younger children (ages 5 to 7). The goal of replicating these studies with younger children was to evaluate whether the same findings for older children and adults would hold for younger children. Though the studies were previously conducted with children as young as 7 years old, evidence from developmental psychology literature prompted us to include even younger participants: typically, as individuals mature from early childhood to adulthood, their cognitive and physical abilities also mature [4, 5]. Thus, the inclusion of younger children will allow for the comparison of patterns across all age groups, and support our overall goal of helping mobile application developers create more age-appropriate apps for children vs. adults, or even universally accessible apps.

Original Study Design

We have previously conducted three studies with adults (over 18 years old) and children (ages 7 to 17) [6, 7, 8] to investigate mobile device input and interaction differences between adults and children. The applications we used were designed specifically for these studies. Each participant completed a gesture task and target task. For the gesture tasks, participants used their finger to draw gestures (i.e., letters, numbers, symbols, and shapes) on the device screen. For the target tasks, participants touched square targets on the phone screen. A summary of the tasks from each study is given in Table 1. We also describe each task briefly to highlight the key points.

	Prelim. [6]	**Study 1 [7, 8]**	**Study 2 [8]**	**Replica**
No. Kids (Ages)	8 (7 to 11)	16 (7 to 16)	25 (10 to 17)	7 (5 to 7)
No. Adults (18+)	6	14	16	N/A
Target Task	Mini Target Task	Target Task	Target Task	Target Task
Gesture Task	No FB Gesture	FB Gesture	No FB & FB Gesture	No FB & FB Gesture

Table 1. Tasks and Studies.

Mini Target Task [6]

Square targets (43 in all) of four different sizes, large (26.4mm), medium (15.8mm), small (10.5mm) and very small (5.29mm), were displayed to the user one at a time. As the participant attempted to touch a target, the application logged the touch event. Participants were allowed one attempt per target only; touches were scored as *hits* or *misses*.

Target Task [7, 8]
The full target task used 104 targets of 4 different sizes: very small (3.175 mm), small (6.35 mm), medium (9.5 mm), and large (12.7 mm), in 13 different interface positions. This task incorporated *edge padding* for half the targets, which caused them to appear close to, but not on, the edge of the screen. The order of targets was designed to evenly represent all possible transitions between target positions and sizes, and no two consecutive targets had the same size or position. Unlike the mini target task, to advance to the next target, the participants had to successfully touch within the boundaries of the visible target. Therefore, multiple attempts for the same target were possible; touches were again scored as *hits* or *misses*.

Gesture Task – Feedback [7, 8]
Participants were shown a screen with text indicating which gesture to make and a "Done" button. Users used their finger to draw gestures on the device screen and press "Done" when finished. The complete gesture set (20 in all) included letters (A, E, K, Q, and X), numbers (2, 4, 5, 7, and 8), symbols (line, plus, arch, arrowhead, and checkmark), and geometric shapes (circle, square/rectangle, triangle, diamond, and heart). Participants were given a paper sheet showing what each gesture should look like, in case they were not familiar with every symbol by name (especially relevant for children). Participants entered an example of each gesture type one after another, and repeated this five times, yielding a total of six examples of each gesture type. As participants drew each gesture, a trace appeared under their finger of the gesture, but they were not able to edit their gestures.

Gesture Task – No Feedback [6, 8]
The no feedback gesture task was identical to the feedback task except participants did not see a trace of the symbol beneath their finger as they drew.

The Replica

We replicated Study 2 (conducted with older children and adults, see Table 1) using the same task applications: participants in the replicated study completed the Gesture Task – No Feedback, Gesture Task – Feedback, and the Target Task. So far, we have had 7 participants in this replication; three were 5 years old, one was 6 years old, and three were 7 years old. Of these participants, four were females, one participant was left-handed, and most self-rated their familiarity with touch input devices to be "average."

Successes of Replica

The primary aspect of the protocol from the original study that was successful was the Target Task: in general, the 5 to 7 year olds were able to complete the Target Task without much difficulty. We believe this was because this task is very short and takes little time (about 1 to 2 minutes) compared to what is required to complete the six iterations of the gesture task (about 8 to 10 minutes). Furthermore, the Target Task required participants to perform an action (touching the interface) with which most children were familiar. In contrast, most of the children were not familiar with all of the gestures they had to draw in the Gesture Tasks and had to practice creating the gestures.

Limitations of the Replication

While the Target Task was a success, we encountered problems with the younger participants not completing all repetitions in the Gesture Tasks. Only 2 of 7 children

completed all iterations of the Feedback and No Feedback Gesture Tasks. The average number of rounds completed was less than 3 for the other children. With the majority completing so little of the task, we did not have enough data to be confident in results from a gesture recognizer (which needs enough data for both a training set and a testing set).

Comparison of Results from the Replica

Target Task – Misses

Table 2 shows the proportion of targets missed on the first attempt in the Target Task for all prior studies [6, 7, 8] and the replica with younger children. The 34% miss rate for the replica is higher than the Study 1 [7, 8] and Study 2 [8] miss rate, which we hypothesize is due to the younger age of the participants (the preliminary study [6] had a higher miss rate because the task only allowed one attempt per target).

	Adults	**Children**
Prelim. [6]	32%	46%
Study 1 [7, 8]	17%	23%
Study 2 [8]	16%	23%
Replica	N/A	34%

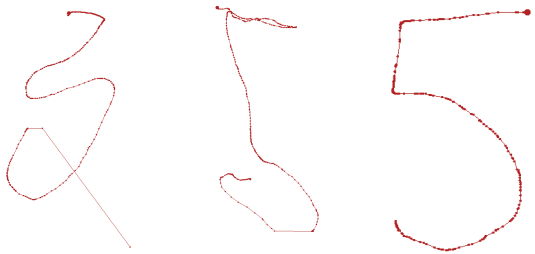
Table 2. Target Task miss results for all studies.

	Adults	**Children**
Study 1 [7, 8]	90% (FB only)	81% (FB only)
Study 2 [8]	91% (FB), 91% (no FB)	82% (FB), 85% (no FB)
Replica	N/A	46% (FB), 49% (no FB)

Table 3. Gesture Task recognition results for three studies.

Gesture Task

Table 3 includes the average per-user recognition results (computed for the replica using the open-source $N multistroke recognizer [9], as in prior work [7, 8]) for both Gesture Tasks across three of the studies. Both in spite of, and as a result of, the lower number of gesture samples collected so far during the replica, the replicated study recognition results are consistent with the overall trend we have found in our work that recognition rates are lower for younger participants. To ensure this finding is robust, we intend to explore ways to encourage children to complete the tasks so that we can examine this trend in more depth for the youngest children. We also hypothesize that the lower recognition rates may be attributed to the grade level of some of the participants (some 5 and 6 year olds had not completed first grade). Children who had been to school had more practice with handwriting and made gestures that appeared be more canonical (Figure 1).



(a) (b) (c)

Figure 1. The gesture for the symbol '5' generated by a five (a), six (b), and seven (c) year old in the replica.

Reasons for Accounted Differences

In general, we found that the results of the replica were consistent with the original studies. However, we have identified four challenges areas with respect to younger

children not completing the gesture task portion of the study that could be useful for doing similar empirical replications with younger children in the future.

Motivation. All participants, adults and children, were compensated $10 for their participation in Study 1, Study 2, and the replica [7, 8] (the preliminary study had no compensation [6]). Though financial compensation may motivate adults, we noted that the delayed financial compensation (receiving $10 after the study vs. immediate rewards throughout the study) might not have been enough motivation for the young children in the replica.

Attention Span. We also noted that the young participants of the replica seemed less focused than the older participants from the original studies [6, 7, 8]. For example, they frequently told stories to the experimenter while completing the tasks, especially during the Gesture Tasks, and many asked for water or breaks during the session. Older participants in the original work did not exhibit this behavior [6, 7, 8].

Research Setting. All of the studies were completed in an academic usability lab with no windows [6, 7, 8]. This setting may not have been inviting and comfortable for the young participants of the replica. In the future, we plan to conduct studies in a more kid-friendly environment, such as a bright room with natural light and pleasant surroundings.

Acknowledgements

This work was partially supported by Department of Education HBGI Grant Award #P031B09020 and National Science Foundation Grant Awards #IIS-1218395 / IIS-1218664. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect these agencies' views.

References

[1] Punch, S. (2002). Research With Children: The Same or Different From Research With Adults? *Childhood 9*(3), 321-341.

[2] Morrow, V. and Richards, M. (1996). The Ethics of Social Research with Children: An Overview. *Children & Society*, 10(2), 90-105.

[3] Druin, A. (1999). Cooperative Inquiry: Developing New Technologies for Children with Children. *Proc. ACM CHI 1999*, 592-599.

[4] Thomas, J.R. (1980). Acquisition of Motor Skills: Information Processing Differences Between Children and Adults. *Research Quarterly for Exercise and Sport 51*(1)*,* 158–73.

[5] Piaget, J. (1983). Piaget's Theory. In P. Mussen, ed., *Handbook of Child Psychology.* Wiley & Sons, New York, NY, USA.

[6] Brown, Q. and Anthony, L. (2012). Toward Comparing the Touchscreen Interaction Patterns of Kids and Adults. *ACM CHI 2012 EIST workshop*, 4pp.

[7] Anthony, L., Brown, Q., Nias, J., Tate, B., and Mohan, S. (2012). Interaction and Recognition Challenges in Interpreting Children's Touch and Gesture Input on Mobile Devices. *Proc. ACM ITS 2012*, 225-234.

[8] Anthony, L., Brown, Q., Tate, B., Nias, J., Brewer, R., and Irwin, G. (In press). Designing Smarter Touch-Based Interfaces for Educational Contexts. *Journal of Personal and Ubiquitous Computing: Special Issue on Educational Interfaces, Software, and Technology*, to appear.

[9] Anthony, L. and Wobbrock, J. O. (2012). $N-Protractor: A Fast and Accurate Multistroke Recognizer. In *Proc. Graphics Interface 2012*, 117-120.