As Good as the Real Thing?

A Comparison of Learning from an Educational Touchscreen App versus a Hands-on Material

Sierra Eisen and Angeline S. Lillard

Corresponding author:
Sierra Eisen
Department of Psychology, University of Virginia
P.O. Box 400400, Charlottesville, VA 22904
(434) 982-5368
sle3jt@virginia.edu

Acknowledgements: Preparation of this article was supported by an American Montessori Society Grant to SE and ASL and grants from the Brady Education and Sir John Templeton Foundations to ASL. During this research, SE was a pre-doctoral fellow of the International Max Planck Research School on the Life Course. We thank the families who participated and Begüm Köseler, Ashley Beamish, Chiara Martignetti, and Emily Wright for assistance with data collection.

Abstract

Children's lives are increasingly entwined with mobile technology, particularly within the educational realm. As touchscreen devices have risen in popularity and accessibility, educational applications (apps) for children have flooded the market (Shuler, Levine, & Ree, 2012). Yet educational mobile technology lacks rigorous comparison between apps and the materials they were designed to replicate. In three studies, we examined geography learning—specifically, the states of Australia—from an educational app and its puzzle equivalent. In Study 1, 32 five-year-olds were randomly assigned to either the puzzle or the app condition. After 20 minutes of interaction with the learning tool, children in the puzzle condition learned significantly more states than did children in the app condition. In Study 2, 32 five- and six-year-olds underwent the same procedure as in Study 1, but then were given the learning tool to use at home for one week. Although children in the puzzle condition initially learned more states, there were no significant learning differences after one week. Children in the app condition used the learning tool during the week for over twice as long (M = 78.75, SD = 65.80) as children in the puzzle condition (M = 32.70, SD = 32.83), and time spent with the learning tool was only correlated with increased learning for the puzzle condition. In Study 3, 32 five- and six-year-olds were run in a social app condition in which the experimenter taught the lesson from the puzzle condition using the educational app. Children in the social app condition learned more states than children from the app condition and marginally more than children from the puzzle condition of Studies 1 and 2. This research provides a crucial investigation of the differences between learning from apps and traditional, hands-on materials and suggests that learning from a touchscreen app is most successful when supplemented with in-person social interaction.

Keywords: learning, touchscreen devices, educational apps

As Good as the Real Thing?

A Comparison of Learning from an Educational Touchscreen App versus a Hands-on Material

The swift rise of touchscreen technology has transformed children's daily interactions with media. Media engagement that was once passive and one-sided has become flexible and interactive. Touchscreen applications or "apps" can respond contingently to user actions, offer rewards for correct responses, give hints after incorrect ones, and adjust difficulty based on user performance. Unlike traditional computers, touchscreen devices are easy to manipulate using simple manual "gestures," which has opened the door for widespread use even among young children. In a 2013 Common Sense Media report, 75% of children under the age of 8 had access to a touchscreen device, and 38% of children under the age of 2 had used a touchscreen before (Rideout, 2013). The exceptional usability of touchscreens has led to an abundance of apps that are geared towards teaching children everything from the alphabet to advanced biology. In fact, more than 80% of the top-selling apps in the Education category of the Apple App Store are aimed at children and teenagers (Shuler, 2012).

To successfully learn real-world-relevant information from a touchscreen device, children must be able to transfer information from an artificial 2D source to the 3D world. Extensive research has shown that infants and young children often fail to transfer information from another common screen—the television—to the real world (Barr & Hayne, 1999; DeLoache et al., 2010; Krcmar, Grela, & Lin, 2007; Roseberry, Hirsh-Pasek, Parish-Morris, & Golinkoff, 2009), a problem referred to as the *video deficit* (Anderson & Pempek, 2005). Socially contingent interactions are associated with better learning (DeLoache et al., 2010). Electronic devices rarely promote social interactivity, and in fact might reduce it: parents' child-directed language is reduced when using an electronic shape sorter as opposed to a physical one (Zosh, Verdine, Filipowicz, Golinkoff, M., Hirsh-Pasek, & Newcombe, 2015). Parental use of mobile devices during mealtimes with their children has been shown to negatively impact parent-child interactions (Radesky et al., 2014). And although parents report frequently reading books or watching TV with their children, they are much less likely to co-use tablets and smartphones (Connell, Lauricella, & Wartella, 2015).

Although not socially contingent, touchscreens (unlike television) can provide contingent feedback through physical interactions with a touch-sensitive screen (Troseth, Russo, & Strouse, 2016). Children's ability to capitalize on touchscreen contingency and transfer learned information from touchscreens to the real world has only recently been studied, and early findings suggest that age plays an important role. Infants and young preschoolers have difficulty transferring learned information, such as novel words and object configurations, between 2D touchscreens and 3D objects (Moser et al., 2015; Zack, Barr, Gerhardstein, Dickerson, & Meltzoff, 2009; Zack, Gerhardstein, Meltzoff, & Barr, 2013). Yet in three recent studies, older children learned from touchscreen apps (Berkowitz et al., 2015), were capable of transferring bi-directionally across 2D and 3D objects (Huber et al., 2015), and even learned equally well from an app and live instruction (Kwok et al., 2016). Although these results are compelling, each study has important and limiting flaws.

Berkowitz and colleagues (2015) explored whether 1st graders' math achievement could be increased with regular use of a math app. Parents and children engaged with the math app at bedtime several times a week. They found that the more families used the app, the better their math achievement at the end of the year, controlling for beginning of year math skills—particularly when the parents were anxious about math. However, their control group used a reading app, rather than hands-on math materials. The lack of a direct comparison between conditions restricts extrapolation of these results and leaves open the question of whether math learning from an app is superior to learning from other educational math materials.

Huber and colleagues (2015) looked at the problem solving abilities of 4- to 6-year-olds using an app and a physical version of the Tower of Hanoi, a puzzle commonly used to assess planning and executive function (e.g., Bull, Espy, & Senn, 2004; Lillard & Peterson, 2011). To solve the puzzle, three disks (small, medium, and large) must be moved from a starting peg to a third peg while maintaining three rules: 1) no more than one disk may be moved at a time, 2) a larger disk cannot be placed on a smaller disk, and 3) the disks must only be placed on pegs. Huber and colleagues found that children became better at the task with practice, regardless of the medium they used to practice. They also found that children who practiced with the app successfully transferred their skill to the physical version. However, the Tower of Hanoi is marketed in the Apple App Store as a problem-solving game similar to checkers or chess, rather than an educational app.

Another recent study examined whether children learn more from face to face instruction or interactions with a touchscreen app (Kwok et al., 2016). Four- to 8-year-olds were introduced to four novel animal facts by either a live female experimenter or by a touchscreen app designed by the researchers that featured a talking cartoon llama. They found that children learned equally well from either a live person or a touchscreen app, with an average two facts learned across either condition. Although this study provides a tightly controlled comparison of learning from direct instruction versus an app, the use of a non-commercial app and the fairly limited amount of information that children could learn limits the impact of the study. Hence none of the prior research strictly compares learning from an educational app to learning from a comparable physical material in a way that approximates children's real-life experiences with educational touchscreen media.

Apps and the devices that support them are attractive to many children. If children enjoy using apps more than traditional tools, this may enhance their ability to learn from apps, as enjoyment and interest are known to enhance learning (Lillard, 2011). However, some of the same elements that might be enticing for children could also be distracting: sound effects, animations, and other non-essential features. Even for traditional storybooks there is evidence that exciting but irrelevant features such as pop-ups detract from children's ability to learn (Chiong & DeLoache, 2012; Tare, Chiong, Ganea, & DeLoache, 2010). In an exploration of e-books, Parish-Morris and colleagues (2013) found that "bells and whistles" detracted from children understanding and remembering a story.

Although the research comparing apps to traditional learning tools has grown in recent years, none of the previously mentioned studies compared commercially available educational apps to a closely matched traditional learning tool. And yet, educational apps and touchscreen devices are increasingly integrated into classrooms to supplement direct instruction from teachers (Richtel, 2011). The present research compares learning from a physical geography puzzle of Australia, presented with a standard teacher lesson, to *iWorldGeography Australia,* an app version of the same puzzle. Children learned about the states and territories of Australia, as well as the island countries surrounding it (hereafter "states" for simplicity), from either the puzzle or the app. The puzzle and its lesson are part of the Montessori preschool education curriculum and the app is designed to closely match this part of the curriculum. We chose these materials due to their similarity to each other and the app's availability within the Education category of the Apple App Store. Importantly, we also expected that children would be unfamiliar with both the tools themselves and the states of Australia, since all participants were from the United States. This allowed us to test their learning without influence from prior knowledge. Therefore, these materials allowed for a tightly controlled comparison of learning from touchscreen apps and learning from traditional physical tools.

These studies addressed three main questions which remain open-ended. First, does children's ability to learn from an app differ from their ability to learn from traditional, physical materials? In the first two studies, we allowed children to interact with the app on their own to imitate how they commonly interact with these tools. We wish to test the assumption that apps can educate children without additional scaffolding. In the final study, we examined instructor-guided learning from apps. Second, if learning differences do exist between these materials, are they due to children's engagement with each? The motivation for placing new technology in classrooms is often centered around engaging children in learning (Richtel, 2011). We wish to know whether the app will be more or less engaging than the puzzle and whether this impacts learning outcomes. Third, does prior experience with touchscreen devices and apps improve learning? We know of no research that has addressed the relationship between app experience and learning. However, because familiarity is generally known to improve learning, one would expect more experienced app-users to learn more from apps.

Study 1

Participants

Thirty-two 5-year-olds (M = 65.10 months, SD = 3.81 months, range = 59 - 72 months; 16 female) were recruited from a database of local families willing to bring their children to the laboratory for research. Children were randomly assigned to either the puzzle condition (n = 16) or the app condition (n = 16). Five additional children were excluded due to inability to complete the experiment (n = 5, 3 from the puzzle and 2 from the app condition). Children were predominantly Caucasian and middle class, reflecting the community from which they were recruited. None of the children were from Australia or the Pacific Rim countries.

Materials

Children in the puzzle condition interacted with a puzzle map of Australia measuring 58.5 x 47 centimeters and constructed from wood. The puzzle contained nine pieces representing the six states of Australia (Western Australia, the Northern Territory, South Australia, Queensland, New South Wales, and Victoria) and three surrounding islands (Tasmania, Papua New Guinea, and New Zealand). The puzzle pieces varied in color and were inset in a blue background representing the South Pacific Ocean.

Children in the app condition interacted with an iPad application called *iWorldGeography Australia* using an iPad mini. The app contained a digital map of Australia that was identical to the puzzle map. The app had seven activity sections, including a blank map of Australia onto which puzzle pieces could be moved, a color map of Australia that recited the name of each state as it was pressed, and a testing section, where children could press a button, hear the name of a state, and then touch the correct state. If the child chose incorrectly, the app would indicate the correct answer by lighting up the correct state and repeating its name.

Procedure

Children were first familiarized with the puzzle or app with a short warm-up where the experimenter introduced Australia and explained it had different parts that they would be learning about. Next, children underwent the procedure for either the puzzle condition or the app condition.

Puzzle Condition. This condition was modeled on a standard Montessori geography lesson. To break the lesson into manageable sections, children were presented with Australian states in three groups of three, for a total of nine states. The experimenter began by pointing to a state, verbally labeling it, and tracing its shape with a finger. Children were then asked to repeat the state name and trace it with a finger. This process was repeated for each of the three states in turn. After each state had been introduced in this manner, the experimenter took the three states out of the puzzle and placed them in front of children. The experimenter asked children to hand her a state, place a state in its proper location on the map, or say the name of a state. This was repeated for each state in counterbalanced order. Finally, the experimenter held up each puzzle piece individually and prompted children to identify the state. This entire process was repeated for all states of Australia. After completing all three rounds, which took about 10 minutes, children engaged in free play with the puzzle for a further 10 minutes by themselves. After the free play period, the experimenter tested children on their ability to identify each state (recognition) and recall each state's name using the physical puzzle. Children were given partial credit for remembering most of a state's name, for example, saying "New Wales" instead of New South Wales.

App Condition. The app condition was designed to determine whether the way children might ordinarily encounter an app, without parent or teacher scaffolding, is as useful for learning as direct instruction. The experimenter presented the seven sections of the app to children by demonstrating the function of each section. Children then engaged with the app by themselves for 20 minutes. After children interacted with the app alone, the experimenter tested children on their recognition and recall for each state using the app. Children were given partial credit for remembering most of a state's name. The sound for the app was turned on throughout the interaction period but turned off for the testing period.

Parent Questionnaire. Parents completed a questionnaire about their children's use of touchscreen devices and educational apps. No use or use less than once a week was categorized as low usage, weekly use was categorized as medium usage, and daily use was categorized as high usage.

Results

Results for each test and condition are shown in Table 1. As would be expected, recognition exceeded recall, as indicated by within-subject (paired samples) *t*-tests: puzzle, $t(15) = 6.27$, $p < .001$, $d = 1.62$; app , $t(15) = 5.22$, $p < .001$, $d = 1.30$. Because recognition and recall test scores were highly correlated, $r(30) = .75$, $p < .001$, they were summed to create a composite memory score. An independent-samples *t*-test was conducted to compare overall memory for Australia's state names. There was a significant difference in the scores for puzzle ($M = 9.28$, $SD = 3.02$) and app ($M = 5.22$, $SD = 4.48$) conditions, $t(30) = 3.01$, $p = .005$, $d = 1.06$. Children in the puzzle condition learned more of Australia's state names than did children in the app condition.

--Insert Table 1 here--

Parental report allowed us to examine the effects of children's level of experience with touchscreens on their performance. We conducted a linear regression to measure the effect of touchscreen usage on learning for children in the app condition and found no relation between touchscreen usage and test scores. Children who were low in their touchscreen use ($n = 3$) were no different in their learning of Australia's state names than children who were considered medium ($n = 9$) or high ($n = 4$) in their use of touchscreen devices.

Discussion

In Study 1, children in the puzzle condition learned more of Australia's state names than children in the app condition on an immediate post-test. We found no relation within the app condition between children's general use of touchscreens and their performance on these tests. Children who were frequent users of touchscreens did not learn more than children who rarely used touchscreen devices.

Study 2

Study 1 showed that children learned more from an experimenter-provided lesson with a physical material than from an app designed to teach the same material. The promise of such apps is that they can to some degree replace teachers and physical, hands-on materials (Richtel, 2011). In this case the standard lesson with the physical material—albeit just at a single time point and 10 minutes in duration—resulted in significantly better learning than time with a commercially available app designed to replace that learning experience.

However, this study took place in a laboratory setting. It is crucial to determine how children learn from these tools in their daily lives, rather than the confines of a testing room. In Study 2, after an initial procedure and test that were identical to those used in Study 1, children brought home either the puzzle or the app to use on their own for one week. After this week, children returned to the laboratory for a second test of their learning. Children could use the material as frequently or infrequently as they wanted at home, and parents were asked to track their children's use. From this study, we hoped to learn how children engage with and learn from apps and puzzles at home. We predicted that children in the puzzle condition would initially learn more of Australia's state names than children in the app condition, in alignment with our findings from Study 1. We further predicted that children who spent more time with the material and were more engaged at home would learn more of Australia's state names by the second test, regardless of condition.

Participants

Thirty-two 5- and 6-year-olds (M = 66.30 months, SD = 5.93 months, range = 60.5 – 83.1 months; 14 female) participated. Although 6-year-olds were included, the average age was not significantly different from Study 1, $t(62) = -0.97$, $p = .338$. Two additional children were excluded due to experimenter error (n = 1) or inability to complete the study (n = 1). Children were randomly assigned to either the puzzle condition (n = 16) or the app condition (n = 16). Children were predominantly Caucasian and middle class. All children were recruited from the same database of local families willing to bring their children to the laboratory for research.

Materials

The puzzle of Australia and the *iWorldGeography Australia* app were the same as those used in Study 1. Children in the puzzle condition also used a 58 x 47 centimeter laminated map of Australia with each state name labeled.

Procedure

Children participated in two experimental sessions spaced seven days apart. The procedure for session 1 was identical to the procedure for Study 1 and included a learning period and an initial post-test (Time 1). Session 2 consisted of a second post-test (Time 2) that was identical to the session one post-test.

Parent Questionnaire and Usage Diary. The parent questionnaire was the same as in Study 1. Parents were also asked to keep track of their children's daily use of the material in a usage diary. Parents recorded each time their children played with the material, the approximate time spent, and their children's level of engagement with the material. Parents scored engagement level on a Likert scale from unengaged (1) to very engaged (5) each time children used the material, which was then averaged across the week. Parents were instructed to allow their children to freely choose whether or not to use the material throughout the week and to provide no direct encouragement. Parents were discouraged from using the material along with their children, since this study focused on children's independent interactions with the materials at home.

Results

Recognition scores exceeded recall scores, as indicated by within-subject (paired samples) *t*-tests, at Time 1: puzzle, $t(15) = 7.79$, $p < .001$, $d = 1.97$; app , $t(15) = 4.21$, $p = .001$, $d = 1.08$, and Time 2: puzzle, $t(15) = 3.85$, $p = .002$, $d = 0.98$; app , $t(15) = 4.78$, $p < .001$, $d = 1.21$. Recognition and recall test scores (shown in Table 2) were again highly correlated at both Time 1, $r(30) = .86$, $p < .001$, and Time 2, $r(30) = .88$, $p < .001$, so recognition and recall were combined into composite memory scores by summing the two scores for each time point.

The first analyses examined whether results of Study 1 replicated. They did: Children in the puzzle condition obtained significantly higher memory scores at Time 1 than did children in the app condition, $t(30) = 2.98$, $p = .006$, $d = 1.06$.

--Insert Table 2 here--

Second, a two-way mixed ANOVA was conducted to examine the effect of condition on children's overall memory across the two testing times. This revealed a main effect of time on children's memory scores, $F(1, 30) = 25.85$, $p < .001$, $\eta p^2 = 0.46$ and a main effect of condition on memory, $F(1, 30) = 6.46$, $p = .016$, $\eta p^2 = 0.18$ (see Figure 1). There was no statistically significant interaction between the condition and the testing time on memory.

--Insert Figure 1 here--

Post hoc analyses were performed using the Bonferroni adjustment. Regarding time, children scored significantly higher at Time 2 (*M* = 12.55, *SE* = 0.82) than at Time 1 (*M* = 8.34, *SE* = 0.77), M_{diff} = 4.20, 95% CI [2.52, 5.89], $p < .001$, suggesting the experience during the intervening week assisted learning. Regarding condition, children in the puzzle condition (*M* = 12.17, *SE* = 0.96) scored higher (overall across the two sessions) than children in the app condition (*M* = 8.72, *SE* = 0.96), M_{diff} = 3.45, 95% CI [0.68, 6.23], *p* = .016.

As in Study 1, parental report allowed us to examine the effects of children's previous experience with touchscreens. Touchscreen usage level was not related to memory scores for children in the app condition during either session. Children who were low in their touchscreen use (*n* = 2) did not differ in their learning of Australia's state names from children who were considered medium (*n* = 7) or high (*n* = 7) touchscreen users.

Parents also recorded their children's daily use of the study materials over one week. An independent-samples *t*-test was conducted to compare the total time that children in the puzzle and app conditions used their respective materials. There was a significant difference between the total usage time for puzzle (*M* = 32.70 minutes, *SD* = 32.83 minutes) and app (*M* = 78.75 minutes, *SD* = 65.80 minutes) conditions, $t(29) = 2.44$, *p* = .021, *d* = 0.93. Children in the app condition used their tool at home over twice as long as children in the puzzle condition. However, there was no significant difference in the frequency with which children used the puzzle (*M* = 4.20, *SD* = 2.68) or the app (*M* = 5.38, *SD* = 2.39), $t(29) = 1.29$, *p* = .207. Instead, the average interaction time with the app (*M* = 13.62 minutes, *SD* = 7.97 minutes) was significantly longer than with the puzzle (*M* = 6.85 minutes, *SD* = 4.74 minutes), $t(29) = 2.85$, *p* = .008, *d* = 1.07, suggesting the app was perhaps more engaging. However, this was not apparent from parent engagement measures: an independent-samples *t*-test indicated no significant difference in parent-rated engagement for puzzle (M = 3.31, SD = 0.68) and app (M = 3.66, SD = 0.89) conditions, $t(29) = 1.22$, $p = .233$.

Correlations were conducted to examine whether time with each material predicted learning, operationalized as the difference in memory score (T2-T1). The predictions were borne out: total time use was correlated with learning, $r(29) = 43$, $p = .015$. Frequency of use was even more strongly related to learning, $r(29) = .60$, $p < .001$. These relations held when considering just the puzzle condition: total time, $r(13) = .74$, $p = .002$; frequency of use, $r(13) = .68$, $p = .006$. The average amount of time children used the puzzle in each interaction was also related to learning, $r(13) = .62$, $p = .014$. By contrast, in the app condition time spent using the app was not related to learning, but frequency of use was marginally related, $r(14) = .47$, $p = .064$. Parent-rated engagement was related to learning for the sample as a whole, $r(29)= .48$, $p = .007$, but not for either condition, although children in the puzzle condition did show a trend, $r(13)= .50$, $p = .057$.

Discussion

Study 2 replicated the results of Study 1: Children in the puzzle condition learned significantly more of Australia's state names at an immediate test session than did children in the app condition. However, after one week during which children had the opportunity to use the materials at home, this degree of advantage was reduced. It is important to note that several children in the puzzle (n = 6) and app (n = 2) conditions of Study 2 received perfect or almost perfect scores by the second session, suggesting a potential ceiling effect.

There were differences between our two conditions in the amount of time children chose to engage with the materials. Overall, children in the app condition used their material over twice as long as did children in the puzzle condition. However, children in the app condition did not use their material more frequently, and they were not more engaged with the material as measured by parent report, than children in the puzzle condition. Yet there were relationships between how much time children engaged with the materials and their learning over the week. For the puzzle condition, time spent with the tool and frequency of use predicted improvement in scores. For the app condition, these factors did not impact learning.

Study 3

Our first two studies indicate that children learn more from their interactions with an experimenter and a physical puzzle than from interacting alone with an app. However, an obvious and intentional confound of these studies is that the puzzle condition involved interactions with a real person who could provide socially contingent responses, while the app condition was limited to the physical contingency of the touchscreen device. Our aim in the previous experiments was to contrast a traditional form of teaching, which involves a live person using physical materials, to a modern form of teaching increasingly promoted in schools and homes, wherein children are handed a touchscreen device with an educational app and expected to learn on their own. We found that even when the touchscreen app is designed to closely match the traditional materials, children learn less from their interactions with a touchscreen.

Recent media recommendations from the American Academy of Pediatrics (AAP) state that children as young as 18 months can use interactive media provided that their parents co-use with them to scaffold learning (American Academy of Pediatrics, 2016). Similarly, the AAP recommends that parents of preschoolers use media devices along with their children to help them understand the content and apply it to the world around them. Currently, parental co-use of touchscreens is infrequent (Connell, Lauricella, & Wartella, 2015; Radesky et al., 2014) but will likely grow as parents are encouraged to engage with their children's touchscreen interactions. In

Study 3, we asked whether preschoolers learn more from an app when an adult gives it a sustained introduction, akin to the lesson of the puzzle condition of Studies 1 and 2.
In this "social app" condition, children used the app along with an experimenter in a structured lesson for about 10 minutes, and then played alone with the app for an additional 10 minutes. We expected children in the social app condition to outperform children in the app condition of Studies 1 and 2. We also expected children in the social app condition to outperform children in the puzzle condition of the two prior studies, namely because the social app condition contained both social interaction within an experimenter-led lesson (as in the puzzle condition) and feedback provided by the app during free play (as in the app condition).

Participants

Thirty-two 5- and 6-year-olds (M = 66.16 months, SD = 5.38 months, range = 57.6 – 76.3 months; 16 female) participated. Although 6-year-olds were included, the average age was not significantly different from Study 1, $t(62) = -0.91$, $p = .367$, or Study 2, $t(62) = 1.02$, $p = .919$. Four additional children were excluded due to familiarity with the study materials ($n = 2$) or inability to complete the study ($n = 2$). All children participated in a social app condition ($n = 32$) and were compared in analyses to the puzzle condition ($n = 32$) and app condition (Time 1 only, $n = 32$) from Studies 1 and 2. Children were predominantly Caucasian and middle class. All children were recruited from the same database of local families willing to bring their children to the laboratory for research.

Materials

The *iWorldGeography Australia* app was the same as in Studies 1 and 2.

Procedure

The procedure for Study 3 was adapted from the puzzle condition procedure for Study 1 and included a learning period, a free play period, and a testing period. Children were presented with the *iWorldGeography Australia* app and the experimenter used the first section of the app, which displayed the full map of Australia, to introduce the name of each state. Since this section of the app recited each state's name as the state was touched, the experimenter kept the sound turned off throughout the lesson. The experimenter presented Australia's states in three groups of three, as in Study 1. The experimenter pointed to each state, verbally labeled it, and traced its outline with a finger, before asking children to label the state and trace its shape as well. This was repeated for each of the three states in turn. After each of the three states had been introduced, the experimenter switched the app to a section where virtual puzzle pieces of each state could be moved onto the map. The experimenter moved the states that were not being taught onto the virtual map, so that only the three target states remained outside of the map. This approximated the set-up of Study 1, where three states were taken out of the puzzle and placed before children. The experimenter then asked children to point to a state, place a state in its proper location on the virtual map, or say the name of a state. This was repeated for each state in counterbalanced order. Finally, the experimenter returned to the first app section that contained a full map and pointed to each state individually, prompting children to identify the state. This entire process was repeated for all states of Australia. After completing all three rounds within about 10 minutes, the sound on the app was turned back on and children were presented with the seven sections of the app. Children engaged in free play with the app for 10 minutes by themselves. After the free play period, the experimenter turned the sound off again and tested children on their recognition and recall of each state using the app. Children were given partial credit for remembering most of a state's name.

Parent Questionnaire. The parent questionnaire was the same as in Studies 1 and 2.

Results

Results for each test and condition are shown in Table 3. As expected, recognition exceeded recall, as indicated by within-subject (paired samples) *t*-tests: puzzle, $t(15) = 6.39$, $p < .001$, $d = 0.63$; app , $t(15) = 5.36$, $p < .001$, $d = 0.36$. Recognition and recall test scores were again highly correlated, $r(30) = .71$, $p < .001$, so they were summed to create a composite memory score.

--Insert Table 3 here--

A one-way ANOVA was conducted to compare overall memory for Australia's state names between the three conditions: puzzle ($n = 32$), app ($n = 32$), and social app ($n = 32$). There was a significant main effect, indicating a difference in the scores for puzzle ($M = 9.97$, $SD = 3.81$), app ($M = 5.66$, $SD = 4.25$), and social app ($M = 12.16$, $SD = 3.74$) conditions, $F(2, 93) = 22.57$, $p < .001$, $\eta^2 = 0.33$ (see Figure 2). Post hoc analyses using Tukey's HSD revealed that children in the social app condition learned significantly more of Australia's state names than children in the app condition, $p < .001$, and learned marginally more of Australia's states than children in the puzzle condition, $p = .073$.

--Insert Figure 2 here--

As in Studies 1 and 2, parental report allowed us to examine the effects of children's previous experience with touchscreens. Touchscreen usage level was not related to memory scores for children in the social app condition. Children who were low in their touchscreen use ($n = 1$) were no different in their learning of Australia's state names than children who were considered medium ($n = 19$) or high ($n = 12$) in their use of touchscreen devices. Although the rates of medium and high touchscreen use in Study 3 appear larger than in the two previous studies, a chi-square test indicated these differences were not significant.

Discussion

In Study 3, children learned more of Australia's states in the social app condition than in the app condition of the two prior studies. They also learned marginally more states than in the puzzle condition of the first two studies. We found no relation within the social app condition between children's general use of touchscreens and their learning.

Why did the social app condition improve learning to this degree? We expect that it is because an adult engaged the child with the material, giving the benefits of social interaction and direct instruction, and then the app subsequently "taught" the child by testing the child and giving feedback after mistakes.

General Discussion

Our primary objective was to directly compare a physical puzzle with a closely-matched commercial app to measure whether children learn more from one or the other tool. In our first study, we found that children who interacted with the puzzle learned more of Australia's state names than did children who interacted with the app. Although it is plausible that children who are more familiar with touchscreen devices would learn more readily from them, we saw no relation between usual touchscreen use and learning for our app condition. Study 2 replicated the findings from Study 1 for the first testing session, with children in the puzzle condition initially learning more than children in the app condition. By the second test, children in the puzzle and app conditions performed comparably but there was a significant main effect for condition, again favoring the puzzle. In Study 3, when children engaged with an app along with a social partner (in this case, the experimenter), they learned significantly more than when they interacted with the app alone.

An important difference between conditions in the first two studies is the level of social interaction that each involved. The limited research on children's co-use of touchscreens suggests that adults and children do not generally use these devices in tandem (Connell, et al., 2015; Radesky et al., 2014). In keeping with this, we chose to initially let the app serve as the teacher. In Studies 1 and 2, the majority of the children's app interactions were without the experimenter, in contrast with our puzzle condition. In Study 3, we assessed whether the differences found in the first two studies were due to a lack of social interaction in the app condition. We found that supplementing the app condition with live social interactions dramatically increased learning from the app. These findings highlight the importance of incorporating social interactivity into preschool children's app use, in support of AAP recommendations (American Academy of Pediatrics, 2016).

Several aspects of these findings will be considered further. One concerns home use of the materials. Study 2 established that children used the app and the puzzle at home an equal number of times and with equal engagement. However, duration of use differed: Children used the app at home twice as long as they used the puzzle. This increased time did not translate into increased learning—children learned the states equally well from the app as from the puzzle after a week with each at home. Interestingly, no relation was found between the additional time spent using the app and subsequent learning. Instead, children in the app condition learned marginally more if they used the app more frequently, suggesting a spaced learning effect. On the other hand, children who used the puzzle did benefit from how much time they spent with the puzzle, as well as how frequently they used it. Time with the puzzle was thus "well spent" in terms of learning.

A second issue of note is the choice of study materials. Unlike many apps, ours did not contain extraneous features such as music or animations. We intentionally chose an app that closely matched the physical puzzle of Australia. Research on superfluous characteristics of apps has suggested they interfere with children's attention and comprehension (Parish-Morris, et al., 2013). Further research should explore the teaching potential of popular apps, with special attention paid to apps that use additional interactive features in ways that promote learning (Hirsh-Pasek et al., 2015). A recent meta-analysis of electronic books (e-books) found that features like simple animations could promote story comprehension, whereas nonessential features like embedded games detracted from comprehension (Takacs, Swart, & Bus, 2015).

How can parents and educators separate the truly educational apps from the vast market of pseudo-educational options? When apps market themselves as educational and are sold in the Educational market of the Apple App Store, many parents likely take the claims at face value. But as Hirsh-Pasek and colleagues (2015) discussed in their recent review of educational apps, most of the apps that are currently available are designed without reference to what scientists know about learning. As parents and educators attempt to navigate the "digital Wild West," research on touchscreens and educational apps should help to inform and guide their decisions (Guernsey, Levine, Chiong, & Severns, 2012).

One limitation of the findings is the reliance on parent report for the home period of Study 2. The relations found between parent measures and learning suggest their reports are reliable. However, it is unclear exactly how parents defined their children's "engagement," since engagement might not look the same for every child. An observational study of children's use of apps and puzzles may provide more detailed data regarding how children used the materials during the week at home. This would shed better light on how children's behavior contributed to learning. A second limitation concerns ceiling effects we observed for some children in Study 2. We used materials that would be novel for children in order to eliminate the effects of prior knowledge. Pilot testing and Study 1 indicated that the Australia materials were sufficiently challenging for children of this age, but several children across both conditions learned all of Australia's states by the second session of Study 2. By limiting the amount that children could learn from the tasks, we may have also limited our findings.

In conclusion, here we found that a physical puzzle given with a lesson by a live adult resulted in more learning than an app intended to replicate that experience. Yet when a live adult gave the lesson on the app itself, learning was somewhat (although not significantly) better than even the physical puzzle with the lesson. Although mobile technology bears some resemblance to its media predecessors, the novelty of the touchscreen has revolutionized mobile devices and limited the assumptions we can draw from past media research. Apps fall somewhere in between the usual dichotomy of passive media viewing and active object use. App users are actively involved in manipulating virtual material with their hands, but the objects are pixel representations rather than solid parts. Not enough is known about how children's learning changes when using touchscreen technology, despite the increasing integration of touchscreens into homes and classrooms. The degree to which learning from apps is like learning from actual materials, and the role of social interaction in app learning, is therefore of deep practical and theoretical importance for both education and cognitive development.

References

American Academy of Pediatrics. (2016). Media and young minds. *Pediatrics, 138*(5), e20162591. doi:10.1542/peds.2016-2591

Anderson, D. R., & Pempek, T. A. (2005). Television and very young children. *American Behavioral Scientist*, *48*, 505–522. doi:10.1177/0002764204271506

Barr, R., & Hayne, H. (1999). Developmental changes in imitation from television during infancy. *Child Development*, *70*(5), 1067-1081. doi:10.1111/1467-8624.00079

Berkowitz, T., Schaeffer, M. W., Maloney, E. A., Peterson, L., Gregor, C., Levine, S. C., & Beilock, S. L. (2015). Math at home adds up to achievement in school. *Science*, *350*(6257), 196-198. doi:10.1126/science.aac7427

Bull, R., Espy, K. A., & Senn, T. E. (2004). A comparison of performance on the Towers of London and Hanoi in young children. *Journal of Child Psychology and Psychiatry*, *45*(4), 743-754. doi:10.1111/j.1469-7610.2004.00268.x

Chiong, C., & DeLoache, J. S. (2012). Learning the ABCs: What kinds of picture books facilitate young children's learning? *Journal of Early Childhood Literacy*, *13*, 225–241. doi:10.1177/1468798411430091

Connell, S. L., Lauricella, A. R., & Wartella, E. (2015). Parental co-use of media technology with their young children in the USA. *Journal of Children and Media*, *9*(1), 5-21. doi:10.1080/17482798.2015.997440

DeLoache, J. S., Chiong, C., Sherman, K., Islam, N., Vanderborght, M., Troseth, G. L., … & O'Doherty, K. (2010). Do babies learn from baby media? *Psychological Science, 21,* 1570-1574. doi:10.1177/0956797610384145

Guernsey, L., Levine, M., Chiong, C., & Severns, M. (2012). *Pioneering literacy in the digital* *wild west: Empowering parents and educators*. New York, NY: The Joan Ganz Cooney Center at Sesame Workshop.

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting education in "educational" apps: Lessons from the science of learning. *Psychological Science in the Public Interest*, *16*(1), 3-34. doi:10.1177/1529100615569721

Huber, B., Tarasuik, J., Antoniou, M. N., Garrett, C., Bowe, S. J., Kaufman, J., & Team, T. S. B. (2015). Young children's transfer of learning from a touchscreen device. *Computers in Human Behavior*, *56*, 56-64. doi:10.1016/j.chb.2015.11.010

Krcmar, M., Grela, B., & Lin, K. (2007). Can toddlers learn vocabulary from television? An experimental approach. *Media Psychology*, *10*, 41–63. doi:10.108/15213260701300931

Kwok, K., Ghrear, S., Li, V., Haddock, T., Coleman, P., & Birch, S. A. (2016). Children can learn new facts equally well from interactive media versus face to face instruction. *Frontiers in Psychology*, *7*. doi:10.3389/fpsyg.2016.01603

Lillard, A. S. (2011). *Montessori: The science behind the genius* (2nd ed.). New York: Oxford University Press.

Lillard, A. S., & Peterson, J. (2011). The immediate impact of different types of television on young children's executive function. *Pediatrics*, *128*(4), 644-649. doi:10.1542/peds.2010-1919

Moser, A., Zimmermann, L., Dickerson, K., Grenell, A., Barr, R., & Gerhardstein, P. (2015). They can interact, but can they learn? Toddlers' transfer learning from touchscreens and television. *Journal of Experimental Child Psychology*, *137*, 137-155. doi:10.1016/j.jecp.2015.04.002

Parish-Morris, J., Mahajan, N., Hirsh-Pasek, K., Golinkoff, R. M., & Collins, M. F. (2013). Once upon a time: Parent-child dialogue and storybook reading in the electronic era. *Mind, Brain, and Education*, *7*, 200–211. doi:10.1111/mbe.12028

Radesky, J. S., Kistin, C. J., Zuckerman, B., Nitzberg, K., Gross, J., Kaplan-Sanoff, M., ... & Silverstein, M. (2014). Patterns of mobile device use by caregivers and children during meals in fast food restaurants. *Pediatrics*, *133*(4), e843-e849. doi:10.1542/peds.2013-3703

Richtel, M. (2011, September 3). In classroom of future, stagnant scores. *The New York Times*, p. A1. Retrieved from http://www.nytimes.com/2011/09/04/technology/technology-in-schools-faces-questions-on-value.html?pagewanted=all&_r=0

Roseberry, S., Hirsh-Pasek, K., Parish-Morris, J., & Golinkoff, R. M. (2009). Live action: Can young children learn verbs from video? *Child Development*, *80*(5), 1360-1375. doi:10.1111/j.1467-8624.2009.01338.x

Shuler, C. (2012). *iLearn II: An analysis of the education category of Apple's app store.* New York, NY: The Joan Ganz Cooney Center at Sesame Workshop.

Takacs, Z. K., Swart, E. K., & Bus, A. G. (2015). Benefits and pitfalls of multimedia and interactive features in technology-enhanced storybooks a meta-analysis. *Review of Educational Research*, *85*(4), 698-739. doi:10.3102/0034654314566989

Tare, M., Chiong, C., Ganea, P., & DeLoache, J. (2010). Less is more: How manipulative features affect children's learning from picture books. *Journal of Applied Developmental Psychology*, *31*, 395–400. doi:10.1016/j.appdev.2010.06.005

Troseth, G. L., Russo, C. E., & Strouse, G. A. (2016). What's next for research on young children's interactive media? *Journal of Children and Media*, *10*(1), 54-62.

doi:10.1080/17482798.2015.1123166

Zack, E., Barr, R., Gerhardstein, P., Dickerson, K., & Meltzoff, A. N. (2009). Infant imitation from television using novel touch screen technology. *British Journal of Developmental Psychology*, *27*(1), 13-26. doi:10.1348/026151008X334700

Zack, E., Gerhardstein, P., Meltzoff, A. N., & Barr, R. (2013). 15-month-olds' transfer of learning between touch screen and real-world displays: language cues and cognitive loads. *Scandinavian Journal of Psychology*, *54*(1), 20-25. doi:10.1111/sjop.12001

Zosh, J. M., Verdine, B. N., Filipowicz, A., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2015). Talking shape: Parental language with electronic versus traditional shape sorters. *Mind, Brain, and Education, 9*, 136-144. doi:10.1111/mbe.12082

Table 1

Means (and Standard Deviations) for Each Test in Experiment 1

	Recognition	Recall
Puzzle	6.06 (2.02)	3.22 (1.46)
App	3.50 (2.34)	1.72 (2.35)

Table 2

Means (and Standard Deviations) for Each Test in Experiment 2

		Recognition	Recall
Puzzle	Time 1	6.44 (2.22)	4.19 (2.44)
	Time 2	7.34 (1.97)	6.38 (2.05)
App	Time 1	3.75 (2.08)	2.31 (2.26)
	Time 2	6.50 (2.88)	4.88 (2.60)

Table 3

Means (and Standard Deviations) for Each Test in Experiment 3

	Recognition	Recall	Total Memory
Social App (Study 3)	7.19 (1.82)	4.97 (2.22)	12.16 (3.74)
Puzzle (Studies 1 and 2)	6.25 (2.10)	3.72 (2.00)	9.97 (3.81)
App (Studies 1 and 2)	3.63 (2.18)	2.03 (2.28)	5.66 (4.25)

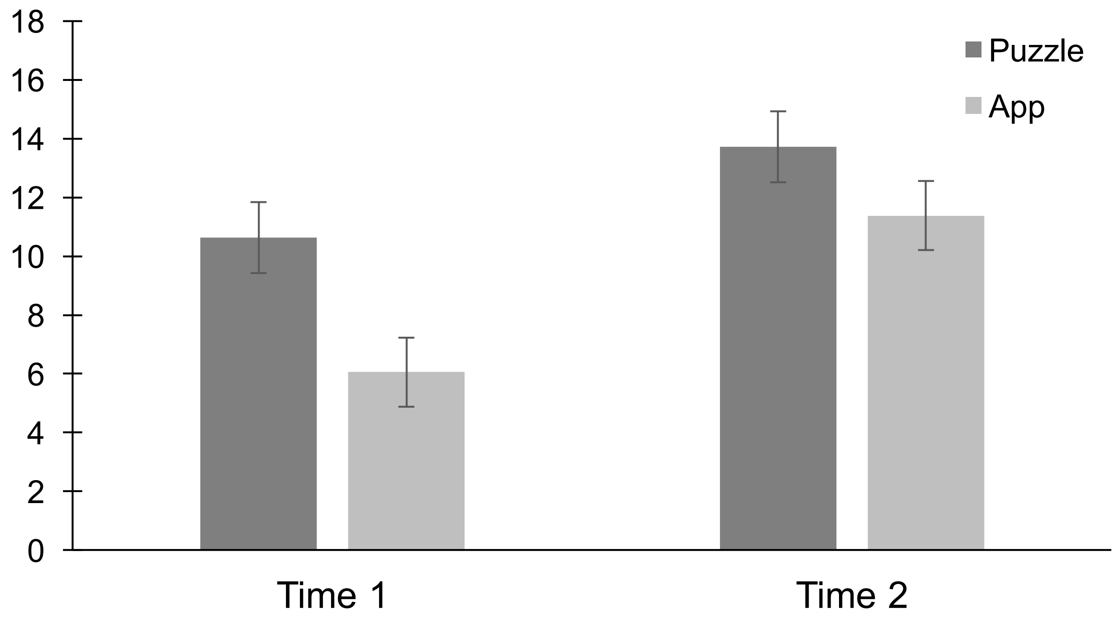

Figure 1. Total memory scores for each condition at first and second testing sessions of Experiment 2

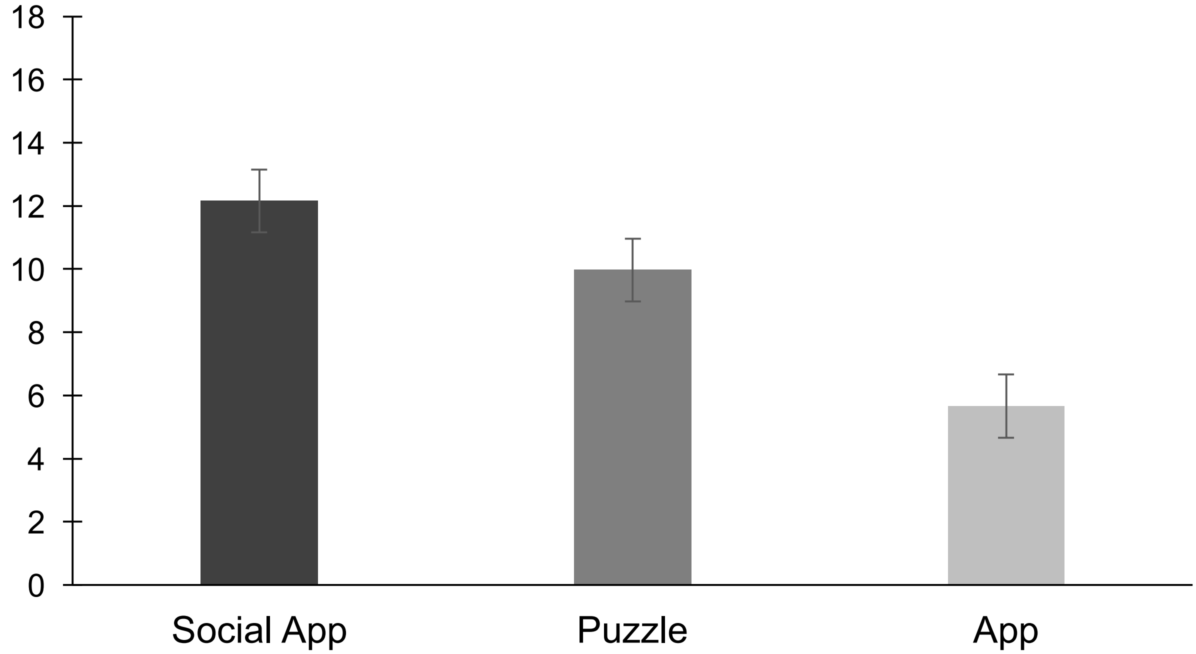

Figure 2. Total memory scores for each condition of Experiment 3