University of Amsterdam

UvA-DARE (Digital Academic Repository)

Haptics and hotspots

creating usable and educational apps for children in the Netherlands

Taylor Piotrowski, J.; Broekman, F.L.

Routledge
Taylor & Francis Group

Haptics and hotspots: creating usable and educational apps for children in the Netherlands

Jessica Taylor Piotrowski [ID][a] and Francette L. Broekman[b]

[a]Amsterdam School of Communication Research ASCoR, Department of Communication Science, University of Amsterdam, Amsterdam, Netherlands; [b]Faculty of Social Sciences, Leiden University, Leiden, Netherlands

ABSTRACT

This study investigated how app design features in educational apps affect app usability (i.e. effectiveness, efficiency, and satisfaction during use) and subsequent learning for Dutch children aged 4–5 years old. Guided by the Capacity Model 2.0 and Cognitive Load Theory, a 2×2 between-subjects experiment was conducted with 128 children (M_{age} = 4.73, SD = .40) to investigate how tactile (i.e. haptic movement: dragging versus tapping) and visual (i.e. hotspots: salient (moving) versus non-salient (non-moving) features in an educational app (M = 4.97 minutes game play) influence app usability and children's learning – namely, receptive vocabulary acquisition. Results lent partial support to study hypotheses. Although children learned nearly five new Dutch words after playing the seeking game only once, the manipulated features did not explain this acquisition. In line with expectations, features did influence usability with salient hotspots proving to be a key predictor of usability. Implications are discussed.

IMPACT SUMMARY

Prior State of Knowledge: Although there is much speculation regarding the impact of app *design* on the usability of and subsequent learning from children's educational apps, empirical knowledge on this topic is lacking.

Novel Contributions: This study is the first empirical investigation to combine predictions of the Capacity Model and Cognitive Load Theory to investigate how educational app design features (haptic and visual) predict app usability and subsequent learning for preschool-aged children.

Practical Implications: For designers, results suggest that the thoughtful use of salient hotspots can lead to improved app usability. But, as demonstrated in the study, usability is not a guarantee for learning. It is crucial to consider *how to* use such features to support learning.

CONTACT Jessica Taylor Piotrowski ✉ j.piotrowski@uva.nl 🖥 Amsterdam School of Communication Research ASCoR, University of Amsterdam, Postbus 15791, 1001 NG, Amsterdam, The Netherlands

App users have roughly 2 million apps at their disposal (Statistica, 2019), with more than 100,000 apps targeting youth. And while exact estimates vary, there is a consensus that children under 6 in industrialized countries around the world use apps for more than an hour per day (Common Sense Media, 2017; Marsh et al., 2015). Unfortunately, the speed of app growth has far outpaced the speed of research and we are left with a comparably limited knowledge base about these apps and their potential effects (Hirsh-Pasek et al., 2015). Thus far, we know that the *children's* app landscape is predominantly comprised of apps which purport to support educational skills, and of these, the majority focus upon literacy development (Chiong & Shuler, 2010; Neumann, 2018; Vaala, Ly, & Levine, 2015).

Perhaps it is not surprising, given the preponderance of children's literacy apps, that most empirical scholarship on app effects has focused on literacy development apps. In the early 2000s, we saw numerous investigations of literacy software and its positive effects on children's emergency literacy skills (de Jong & Bus, 2002, 2004; De Jong & Bus, 2003). With time, we saw increased interest in electronic books and literacy outcomes (Korat & Shamir, 2007; Piotrowski & Krcmar, 2017; Shamir, 2009). These studies highlighted the power of design features, for example, design elements which drew attention to specific words as a key means of supporting learning (Roskos, Brueck, & Widman, 2009). Today, touchscreen technology has far more enhanced design features – with tactile, auditory, and visual elements designed to elicit and engage a user's attention. And while the touchscreen environment holds the promise of superior learning opportunities when compared to earlier technology such as e-books (Kirkorian, Choi, & Pempek, 2016), there is limited research on the effects of tactile-based interfaces on children's literacy development. Moreover, of the studies that do exist, research has not consistently supported the potential of these apps (see Russo-Johnson, Troseth, Duncan, & Mesghina, 2017 for details).

While methodological differences in design or measurement may explain these mixed results, it is also possible that it reflects differences in app design. In particular, it may be (in)efficiencies in app design which explain when and if an educational app is able to reach its curricular goal. Specifically, when the design of an app is mismatched with the child's abilities, educational content comprehension is likely to suffer as a result of poor usability (i.e. the degree to which something is able or fit to be used; Fisch, 2004; Gee, 2006). Yet, our knowledge of the effects of educational app *design* is limited primarily to theoretical speculation (Fisch, 2016). The aim of this study is to address this by investigating how specific app design features affect app usability and subsequent literacy development among 4–5-year-old children. In particular, this study looks at vocabulary development as vocabulary is one of the most crucial precursors of literacy development (Whitehurst & Lonigan, 1998).

Learning from technology

Capacity model 2.0

In 2000, Fisch posited the capacity model (CM) – a model of children's learning from educational television (Fisch, 2000). The model posits that educational television consists of both narrative and educational content and, for the educational content to be

effectively comprehended, the cognitive demands for the narrative cannot exceed the demands necessary for processing the educational content. Yet, this model was not developed with digital (interactive) media in mind. In response, Fisch has recently offered what he refers to as *Capacity Model 2.0* – an extension of the model to educational games (Fisch, 2016). Like its predecessor, the model pays careful attention to the relationship between narrative and educational content. However, the 2.0 version also notes that the gameplay influences working memory allocation. Specifically, high gameplay processing is said to leave fewer resources available for narrative and educational content processing (Fisch, 2016). Moreover, the model posits that a high integration between gameplay elements and educational content reduces the degree to which they must compete for cognitive resources.

Similar to narrative processing in the Capacity Model 1.0, gameplay processing can never be abandoned in favor of the educational content. As such, it is crucial that educational content is successfully embedded within a game so that the gaming elements work to support the educational content rather than drain cognitive resources. In other words, design is crucial. Design elements must match the user's abilities if one hopes to maximize available cognitive resources and support learning. This practice of creating educational experiences that fits the user's abilities is often referred to as *Instructional Design* (Gagne & Briggs, 1974; Reigeluth & Carr-Chellman, 2009) and is prominent in children's app design.

Instructional design

Instructional Design (ID) focuses on designing products that lead to an efficient, effective, appealing, engaging, and inspiring acquisition of knowledge (Merrill, Drake, Lacy, Pratt, & Group, 1996). One of the most influential theories guiding ID, also aligned with Fisch's Capacity Model, is Cognitive Load Theory (CLT). CLT's basic assertion is that any design must take into account working memory in order to prevent overload and subsequent learning deterioration (Paas, Tuovinen, Tabbers, & Van Gerven, 2003).

In CLT, cognitive load is classified as intrinsic, extraneous, and germane cognitive load (Hollender, Hofmann, Deneke, & Schmitz, 2010). Intrinsic cognitive load is related to the complexity of the learning task and is directly related to the expertise of the learner. For example, in developing a vocabulary app, the complexity of the target vocabulary and current vocabulary skills of the learner will together lead to intrinsic load. Extraneous cognitive load is associated with inappropriate teaching of the material or other activities that are irrelevant to the learning task. Here, in a game space, this might include game tasks that are irrelevant to learning the target vocabulary. Finally, germane cognitive load results from the acquisition of new schemas and is considered beneficial for learning. Intrinsic cognitive load can provide the potential for germane cognitive load, and task complexity may cause extraneous cognitive load (Schnotz & Kürschner, 2007).

CLT breaks cognitive load into two subcomponents: mental load (a priori estimate of the expected cognitive capacity demands) and mental effort (the actual cognitive load allocated). In addition to the content of the task at hand, mental load and mental effort are also dependent upon the usability of the technology (Schnotz & Kürschner, 2007). Increased usability, conceptualized as a combination of *effectiveness* (i.e. accuracy and completeness of achieving goals), *efficiency* (i.e. the resources expended to complete

goals), and the *satisfaction* that a user experiences when using technology (I see ISO 9241-11 in Bevan, Carter, & Harker,2015), reduces draw on cognitive resources. ID focuses on ensuring usable experiences so that available working memory is allocated to the learning goal.

Early studies on technology design with children confirm that usability can be enhanced when particular features are present or absent (Laurillard, 2009). For example, usability is improved when navigation icons are clear, consistent, and intuitive; when clickable hotspots are large enough and set far enough apart for users to click easily on the options they desire; and when these hotspots are signaled in some way so that users know where to click (Piotrowski & Krcmar, 2017; Russo-Johnson et al., 2017). Additionally, particular features can enhance user enjoyment. For example, a clear in-game goal enhanced students' enjoyment (Kiili, De Freitas, Arnab, & Lainema, 2012) while a reward system (e.g. points; collectables) can lead to greater enjoyment for young children (Ronimus, Kujala, Tolvanen, & Lyytinen, 2014). This and similar research, in combination with the CM and CLT, suggests that *design* elements (i.e. features) in educational apps which lead to improved usability should support learning by reducing cognitive load. Consequently, in this study, we investigate the mediating role of usability in the relationship between design features and learning.

Design features

Apps have numerous features which are expected to influence the usability of, and potential learning from, an educational app. Design features can be divided into *visual* and *tactile* dimensions whereby visual features represent the way in which the content is presented and tactile features the possible interactions with the content.[1] In this study, we manipulate two design features: *haptic movements* (a tactile feature) and *hotspots* (a visual feature). Given the prevalence of literacy development children's apps, we have opted to investigate these features within the context of a vocabulary development app.

Haptic movement

Touchscreens allow children to manipulate content in different ways: tapping, swiping, and dragging, and even tipping the device. This ability to recognize the user's actions is called haptic movement (Piotrowski & Krcmar, 2017; Russo-Johnson et al., 2017). When the user provides "input" at a specific point on the screen through tactile movement (e.g. a pinch, flick, double-tap, press, etc.), the technology responds with "output" that belongs to this specific point. Based on CM 2.0 and CLT, one would expect that as haptic movement increases in difficulty, usability will decrease, and learning from the app will be impaired.

Although haptic movements vary, the most prominent haptic movements in children's apps are *tapping* and *dragging* (Nacher, Jaen, Navarro, Catala, & González, 2015). Tapping is a short one-finger touch, often used for selection, while dragging is a continuous one-finger movement, often used for relocation of objects. Research shows that, even though 4-year-olds are able to perform both movements (Nacher et al., 2015), children experience more difficulties with dragging (Aziz, 2013) which may be because dragging is a more complex haptic input that requires greater cognitive capacity (Russo-Johnson et al., 2017).

This might, in part, be explained by the fact that tapping is a relatively gross motor skill when compared to the finer motor skill of dragging and thus tapping may require comparably less cognitive energy (also see Bedford, Saez de Urabain, Cheung, Karmiloff-Smith, & Smith, 2016; Lin, Cherng, & Chen, 2017 for comments on this topic). As such, dragging may be less *usable* than tapping suggesting the following hypothesis:

H1: Tapping leads to improved vocabulary learning when compared to dragging (H1a). This relationship is mediated by usability (H1b) such that tapping enhances usability to a greater extent than dragging (H1c).

Hotspot salience

Just as haptic movements are prominent tactile features in children's apps, hotspots are equally prominent visual design features (Vaala et al., 2015). Hotspots are clickable items whereby when an aspect of the screen is touched, it responds (Piotrowski & Krcmar, 2017). Hotspots can be salient or non-salient. In the case of salient hotspots, the content is typically moving and blinking to draw attention whereas non-salient hotspots are non-moving and non-blinking. Meta-analytic research suggests that hotspots can be distracting for youngsters unless they are directly linked with the educational task (i.e. congruent versus incongruent hotspots; Takacs, Swart, & Bus, 2015; also referred to as relevant or irrelevant). However, when they are congruent or otherwise relevant to the educational task, they can be an asset to the learning experience. For example, in an e-book study, children that heard explanations of difficult words in the form of hotspots outperformed children without hotspots (Smeets, Van Dijken, & Bus, 2014). But, type of hotspot does seem to matter with salient hotspots requiring less cognitive resources than non-salient hotspots (Chiong & DeLoache, 2013). As such, we might expect that salient relevant hotspots are associated with improved usability and subsequently improved learning when compared to non-salient relevant hotspots.

H2: Salient hotspots lead to improved vocabulary learning when compared with non-salient hotspots (H2a). This relationship is mediated by usability (H2b) such that salient hotspots enhance usability to a greater extent than non-salient hotspots (H2c).

Haptic movement & hotspot salience

Of course, multisensory activities that involve *both* visual and tactile features can also facilitate learning (Neumann, 2014). Based on the propositions of the CLT and CM, we expect that the combination of low-cognitive load tactile design features (i.e. tapping haptic movement) and low-cognitive load visual design features (i.e. salient hotspots) will be most usable and lead to the greatest vocabulary learning.

H3: *There is an interaction effect of haptic movement and hotspot salience on children's vocabulary learning such that tapping and salient hotspots have the strongest positive effect on vocabulary learning compared to other conditions (H3a). This relationship is mediated by usability (H3b) such that tapping and salient hotspots together enhance usability to a greater extent than other conditions (H3c).*

Methods

Design and procedures

To test hypotheses, a 2 (haptic movement) x 2 (hotspot salience) between-subjects randomized experiment was conducted. After receiving ethical approval, recruitment and study procedures commenced. All testing was conducted within available spaces at the children's schools with only the child and researcher present – using small available children's desks to work upon. Testing sessions lasted, on average, 30 minutes. At the onset, the researcher asked the child two simple questions to create a comfortable atmosphere (i.e. "What is your favorite color?" and "What is your favorite toy?"). Following this, children completed a verbal comprehension assessment. Children were then randomly assigned to one of four apps for game play (see Stimuli section). The children were told "I have a new game that was made for kids like you. You can play it while I do my homework". The child was then given an iPad with the game open. After a practice round with one sample screen on the iPad using images not included in the testing stimuli, the child played as long as it took him/her to finish the game. On average, children in the study required slightly less than 5-minutes to complete the game play ($M = 298.09$ seconds, SD = 119.64). The researcher pretended to do her homework during this time. If the child asked any questions, the researcher answered with "just try it out a bit more". Following gameplay, children completed assessments to evaluate usability, learning, and working memory. Once all assessments were complete, the researcher told the child that he/she had done a good job and escorted him/her back to the classroom.

Participants

A priori power analyses for a 2×2 between-subjects design with a medium effect size indicated that 128 children were needed to achieve .80 statistical power. Matching this, 128 children (64 boys) were recruited for participation. Informed parental consent was required for all participants. All children were Dutch native speakers ($M_{age} = 4.73$, $SD = .40$). Children were randomly assigned to one of 4 conditions: Non-Salient Drag (n = 32), Salient Drag (n = 31), Non-Salient Tap (n = 32), Salient Tap (n = 33). There were no differences in condition by age or gender.

Stimuli

An iPad (2017 model) was used to display the stimulus material. An app was created for this study which aimed to teach the child the names of ten unfamiliar vegetables, such as "turnip", and "rutabaga" (referred to as *vergeten groenten* – forgotten vegetables – in

Dutch). A pilot test (n = 12; using a convenience sample of children that were not part of the full sample) confirmed that children aged 4–5 were unfamiliar with all of the vegetables used in the stimulus material. The app was produced by a design professional and was stylistically similar to many existing vocabulary apps.[2]

The app began with a practice screen. This took the form of a sample game screen (unrelated to the target words in the comprehension assessment) so that children could see and use the interface. Thereafter, the child was exposed to 10 screens (randomly ordered), consistent with the goal to assess one target vegetable per screen. On every screen, seven vegetables were randomly selected to appear one at a time on a white plate on a table (see Figure 1). A voice dictated the name of each vegetable as it appeared. When all vegetables were displayed, the voice asked "can you put the X on the plate?". If the child did not put the correct vegetable on the plate or if he/she did not respond within 1 minute, the question was repeated. If there was no correct answer or response after 1 additional minute, the app placed the correct vegetable on the plate and moved on to the next screen. This was done to ensure all children would be exposed to all 10 different screens (thus all ten target vocabulary words). In either situation, when the correct vegetable was on the plate the child heard "Yes, very good! The X is on the plate!"

Figure 1. **Stimuli Screenshot:** A screenshot with different vegetables surrounding the plate. A voice dictated the name of each vegetable as they appeared. Then, the voice asked "can you put the X [vegetable name] on the plate?" If in the salient condition, the correct vegetable wiggled (otherwise, no movement). If in the tapping condition, tapping on correct vegetable moved item to place (otherwise, user had to drag vegetable to plate).

Of note, given the importance of repetition for vocabulary learning, repetition was structurally built into the game. Specifically, every child heard the vegetable name at least three times – once when announced, once when moving to plate, and once before moving onwards. In addition, when the child touched the screen, the voice would dictate the word again, increasing the repetitions. To manipulate hotspot salience, the app would either display the correct answer highlighted as well as wiggling (salience) or it had no visual cues after the question was asked (non-salience). To manipulate haptic movement, the child could place the correct vegetable on the plate through either tap or drag input. By crossing these factors, a total of 4 different apps were created: Non-Salient Drag, Salient Drag, Non-Salient Tap, Salient Tap.

Measurements

Mediator – Usability

Usability is a multi-dimensional construct which the International Organization for Standardization (ISO) has defined 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 (Frøkjær, Hertzum, & Hornbæk, 2000).

Usability: Effectiveness. *Effectiveness* is the accuracy and completeness with which users achieve specified goals. The most common metric for effectiveness is completion rate. Completion rate is calculated by dividing the number of successfully completed tasks by the number of tasks undertaken (Sauro & Kindlund, 2005). In this study, log data recorded the number of clicks per screen ($M = 3.56$, $SD = 2.75$). We divided the correct number of clicks to complete the screen successfully (in the tap-conditions twice, in the drag-conditions once) by the number of logged clicks per screen to arrive at a relative success rate per screen. When the child did not touch a screen, the score for that screen was zero. Effectiveness was the sum of all ten screens, ranging from 0 (low) to 10 (high) ($M = 7.52$, $SD = 1.94$).

Usability: efficiency

Efficiency is the resources expended in relation to the accuracy and completeness with which users achieve goals (e.g. time spent). The most commonly used measure for efficiency is time-on-task. Here, efficiency was measured through time-on-task of the child compared to an expert user of the app (i.e. the total time it took a researcher to complete all ten screens) (Bevan, 2006). Log data registered the exact number of seconds that the app was used by the child ($M = 298.09$ seconds, $SD = 119.64$). In addition, log data were used to assess (across three attempts) the mean number of seconds it took an expert user of the app (project researcher) to complete all ten screens (in both Tap conditions 202 seconds, in Drag conditions 199 seconds). Expert time was divided by the time taken by the child to complete the task ($M = 0.73$; $SD = 0.17$), ranging from 0 (low) to a maximum score of 1 (high).

Usability: satisfaction

Satisfaction describes the comfort and acceptability of use. There was, to our knowledge, no existing measure for the usability dimension "satisfaction" for young children. Since previous research has used scales for students or more advanced users, it was necessary to adapt (one of) these measures for a younger audience. The System

Usability Scale (SUS; Brooke, 2013) was determined to be appropriate for adaptation given its length and validity. After altering nine of the original ten items for child appropriateness (one was impossible to translate), the items were piloted with children aged 4–5 (N = 20). Children were unable to answer two items. These were items that concerned thinking about others ("do you think other children will be able to play this game?") or the effort put in to play the game ("did you have to put in effort to play the game?"). Both require more sophisticated cognitive abilities than is typical for this age group, offering a likely explanation as to why these items were found problematic. These items were dropped, resulting in a 7-item scale with questions such as "did you like the game?" and "would you like to play the game more often?" The complete list of tested items can be found on the Author's website. For each question, the child first answered using a simple bivariate response (e.g. "yes, I liked it" or "no, I did not like it"). Following this, a smiley-rating system was used to obtain more variance. For example, if the child responded "yes, I liked it", s/he would be asked "did you like it a bit" or "did you like it a lot?" This resulted in a 4-point response option with higher scores indicating greater satisfaction (M = 3.38, SD = .64, $α$ = .78).

Dependent variable – vocabulary learning

We created a receptive vocabulary task to measure vocabulary learning similar to other researcher-developed measures (Linebarger, Moses, & McMenamin, 2010; Linebarger, Piotrowski, & Vaala, 2007). This assessment was administered after the usability assessment. Using a laptop, children saw ten different screens. For each screen, the researcher asked the child to point out a specific vegetable (e.g. "Can you point out the turnip?"). Each screen had four images of vegetables, of which one was correct. Incorrect answers received zero points, while correct answers received 1 point. Total recognition represented the sum-score of all ten screens, ranging from 0 (no recognition) to a maximum score of 10 (perfect recognition) (M = 4.70, SD = 1.89).

Covariate – verbal comprehension

To obtain an indicator of children's verbal comprehension at study onset, the 33-item Information subtest of the Wechsler Preschool and Primary Scale of Intelligence was administered to all participants. This measure was administered prior to app use (M = 11.64, SD = 2.66).

Covariate – working memory capacity

To obtain an indicator of children's working memory capacity, we implemented the modified children's version of the Missing Scan Test (MST), previously validated in Roman, Pisoni and Kronenberger's work (2014). The "missing scan" methodology consists of presenting the participant with a set of digits (or objects) and then reproducing the same set again in a randomized order with one of the original missing. Then, the participant reports what is missing. Each trial is progressively more difficult than the last. In this set-up, digits were replaced with 30 physical Beanie Babies[tm]. The memory

set size began with three animals and increased in length by one animal each time the child correctly reported the missing item. Working memory capacity was defined as the longest set size of animals that the child could correctly scan. The test concluded when the child failed to correctly name the missing animal on two trials of the same memory set size or correctly completed a set size of 10. As such, a score of 10 indicates the highest working memory capacity ($M = 4.62$, $SD = 1.93$).

Covariate – child characteristics

The school provided information about age ($M = 4.73$, $SD = .40$, Range $= 1.69$) and gender ($0 = $ boy, $1 = $ girl) of the participating child.

Results

Descriptive statistics

Figure 2 presents the conceptual model guiding study analyses. All children were randomly assigned to one of four conditions: Non-Salient Drag (n = 32; 16 boys; $M_{age} = 4.69$, $SD = .43$), Salient Drag (n = 31; 16 boys; $M_{age} = 4.78$, $SD = .43$), Non-Salient Tap (n = 32; 15 boys; $M_{age} = 4.72$, $SD = .38$) and Salient Tap (n = 33; 17 boys; $M_{age} = 4.74$, $SD = .38$). There were no significant differences between conditions in age ($F(3,124) = .28$, p $= .84$) or gender ($\chi^2 (3) = .19$, $p = .980$). Children were, on average, able to effectively complete 7.52 ($SD = 1.94$) out of 10 screens, with .73 relative efficiency ($SD = .17$) and perceived the app to be highly satisfactory ($M = 3.38$ of a maximum of 4, $SD = .64$). Most striking, children were able to learn, on average, 4.7 words ($SD = 1.89$) out of ten. For context, by chance, we would expect children to get 25% of the questions correct, i.e. a score of 2.5 – this is significantly lower score than achieved here ($t(127) = 13.12$, p $< .001$). Correlations are presented in Table 1.
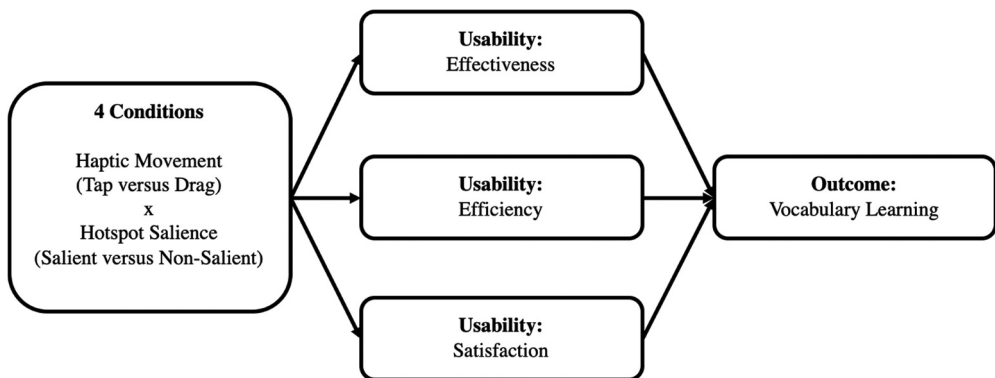
Figure 2. Conceptual Model Guiding Analyses The conceptual model guiding the study and related analyses is depicted in Figure 2. The 2 × 2 design resulted in 4 conditions. Analyses evaluated the extent to which each manipulation independently and jointly predicted vocabulary learning and the extent to which usability (modeled as 3 parallel processes) mediated this process. All analyses controlled for age, verbal comprehension, and working memory.

Haptic movement (H1abc)

Consistent with expectations, in the tapping condition, children learned 4.85 words on average (95% CI [4.40–5.28]) whereas in the dragging condition children learned 4.54 words (95% CI [4.04–5.03]). A bootstrapped regression model (controlling for age, verbal comprehension, and working memory, see Table 2) showed no significant direct relationship between haptic movement and vocabulary learning ($B = -.24$). H1a is rejected.

Although no direct effects were detected, it is possible for indirect relationships to be present without the presence of direct effects (Hayes, 2009). To that end, a multiple mediator model analysis was conducted (Preacher & Hayes, 2008). The model controlled for age, working memory, and verbal comprehension. To address issues of normality, all estimates were bootstrapped. Results indicated that effectiveness, efficiency, and satisfaction were not significant mediators, rejecting H1b. See Table 3.

Finally, to investigate main effects of haptic movement on the usability attributes (H1c), a MANCOVA was conducted with haptic movement as the independent variable and effectiveness, efficiency, and satisfaction as dependent variables (bootstrapped; controlling for age, verbal comprehension, and working memory; see Table 4). Haptic movement had a significant direct effect on effectiveness ($B = 1.08$, SE $_{bootstrapped}$ = .25) and efficiency ($B = .08$, SE $_{bootstrapped}$ = .02). As expected, children in the tapping condition ($M = 8.06$, $SD = 1.98$) completed the tasks more effectively than children in the dragging

Table 1. Correlation matrix for model variables.

	1	2	3	4	5	6	7	8	9
1. Haptic Movement	1								
2. Hotspot Salience	−0.016	1							
3. Usability: Effectiveness	−0.285**	0.645***	1						
4. Usability: Efficiency	0.249**	0.590***	0.647***	1					
5. Usability: Satisfaction	0.050	0.191*	0.319***	0.356***	1				
6. Vocabulary Learning	−0.081	−0.004	0.133	0.041	0.088	1			
7. Age	0.009	0.072	0.150	0.280**	0.192*	0.145	1		
8. Verbal Comprehension	−0.032	0.077	0.162	0.208*	0.196*	0.326***	0.369***	1	
9. Working Memory	−0.048	−0.126	0.088	0.070	0.055	0.263**	0.155	0.216*	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Note: Haptic Condition where 0 = tapping/1 = dragging; Hotspot Condition where 0 = no salience/1 = salience; Age = years; n = 128

Table 2. Regression tables for hypothesis 1a, 2a, 3a.

	Hypothesis 1a		Hypothesis 2a		Hypothesis 3a	
	B	95% CI	B	95% CI	B	95% CI
Haptic Movement	−0.238	[−0.761,0.285]			−0.465	[−1.430,0.499]
Hotspot Salience			−0.003	[−0.619,0.612]	−0.233	[−1.032,0.566]
Age	0.062	[−0.852,0.976]	0.054	[−0.637,0.744]	0.054	[−0.818,0.926]
Verbal Comprehension	0.197*	[0.061,0.333]	0.198*	[0.050,0.347]	0.197*	[0.082,0.926]
Working Memory	0.195*	[0.018,0.372]	0.198*	[0.0230,0.372]	0.192*	[0.013,0.370]
Condition Interaction					0.454	[−0.746,1.654]
_constant	1.329	[−2.844,5.501]	1.222	[−1.852,4.296]	1.497	[−2.417,5.411]
N	128		128		128	
R^2	0.149		0.145		0.153	

* $p < 0.05$
Note: Haptic Condition where 0 = tapping/1 = dragging; Hotspot Condition where 0 = no salience/1 = salience; Age = years; n = 128

Table 3. Multiple mediation pathways for hypothesis 1b, 2b, and 3b.

		Hypothesis 1b			Hypothesis 2b				Hypothesis 3b		
		B	BCa- LL	BCa -UL	*B*	BCa- LL	BCa -UL	Condition	*B*	BCa- LL	BCa -UL
Indirect	Effectiveness	−0.184	−0.483	0.115	0.429	−0.197	1.127	2	−0.100	−0.428	0.050
								3	0.606	−0.300	1.493
								4	0.294	−0.113	0.830
	Efficiency	−0.156	−0.450	0.138	−0.286	−0.776	0.216	2	−0.220	−0.761	0.287
								3	−0.380	−1.186	0.479
								4	−0.424	−1.308	0.529
	Satisfaction	0.004	−0.065	0.073	0.012	−0.103	0.177	2	0.004	−0.082	0.186
								3	0.010	−0.130	0.234
								4	0.012	−0.152	0.248
Direct	Total	0.10	−0.62	0.81	.159	−1.051	0.733	All	−0.75	−.3.50	1.99

* $p < 0.05$

Note: BCa = Bias-Corrected and Accelerated CI (2000 bootstrap); LL = lower level confidence interval; UL = lower level confidence interval; 2 = No Salient, Drag; 3 = Salient, Tap; 4 = Salient, Drag; No Salient-Tap is reference category; n = 128

condition ($M = 6.96$, $SD = 1.74$). However, children in the dragging condition ($M = .78$, $SD = .13$) completed the tasks more efficiently ($M_{tap} = .69$, $SD = .19$). Results provide partial support for H1c.

Hotspot salience (H2abc)

Children in the salient hotspots condition learned 4.69 (95% CI [4.30–5.07]) words on average whereas in the non-salient hotspot condition children learned 4.70 words (95% CI [4.16–5.25]). A bootstrapped regression model (controlling for age, verbal comprehension, see Table 2) showed no significant direct relationship between hotspot salience and vocabulary learning ($B = −.003$). H2a is rejected.

Table 4. MANCOVA results for hypothesis 1c, 2c, 3c.

	Effectiveness	Efficiency	Satisfaction
Hypothesis 1c	B	B	B
Haptic Movement	−1.085*	0.085*	0.67
Age	0.521	0.096*	0.218
VC	0.077	0.008	0.35
Working Memory	0.035	0.001	0.001
_constant	4.527*	0.133	1.898*
Hypothesis 2c	B	B	B
Hotspot Salience	2.530*	0.197*	0.224*
Age	0.279	0.083*	0.203
VC	0.043	0.004	0.031
Working Memory	0.150*	0.008	0.009
_constant	3.746*	0.149	1.902*
Hypothesis 3c	B	B	B
Haptic Movement	−0.499	0.149*	0.106
Hotspot Salience	3.035*	0.258*	0.258
Condition Interaction	−1.064*	−0.120*	−0.067
Age	0.339	0.0824*	0.201
VC	0.036	0.005	0.031
Working Memory	0.144*	0.010	0.011
_constant	3.821*	0.064	1.845*

* $p < 0.05$

Note: VC = verbal comprehension, n = 128

Similar to hypothesis 1, a multiple mediation model was then conducted to evaluate the potential pathway of hotspot salience through usability to vocabulary learning (Preacher & Hayes, 2008). Results indicated that none of the three components of usability were significant mediators, thus rejecting H2b (see Table 3).

Finally, to investigate main effects of hotspot salience on the usability attributes (H2c), a MANCOVA model was conducted whereby hotspot salience served as the independent variable and effectiveness, efficiency, and satisfaction as dependent variables (bootstrapped; controlling for age, verbal comprehension, and working memory; see Table 4). Hotspot salience had a direct effect on all attributes of usability: effectiveness ($B = 2.53$, SE $_{bootstrapped} = .25$), efficiency ($B = .20$, SE $_{bootstrapped} = .02$), and satisfaction ($B = .22$, SE $_{bootstrapped} = .11$). Children in the salient hotspot condition ($M = 8.77$, $SD = 1.40$) completed the tasks more effectively compared to children in the non-salient condition ($M = 6.28$, $SD = 1.57$). They were also more efficient ($M_{salient} = .83$, $SD = .10$; $M_{non-salient} = .63$, $SD = .17$), and more satisfied ($M_{salient} = 3.50$, $SD = .49$; $M_{non-salient} = 3.26$, $SD = .74$). Thus, these results support Hypothesis 2c.

Interaction effect of haptic movement and hotspot salience (H3abc)

A bootstrapped regression model (controlling for age, verbal comprehension, and working memory, see Table 2) indicated that there was no significant interaction effect of haptic movement and hotspot salience on vocabulary learning ($B = .45$), rejecting H3a. As with the analyses for Hypothesis 1b and 2b, a multiple mediation model with bootstrapping was calculated to assess H3b. In this case, condition was a multi-categorical independent variable. Effectiveness, efficiency, and satisfaction again served as mediators and vocabulary learning as dependent variable. The model controlled for age, verbal comprehension, and working memory. None of the attributes of usability mediated the relationship. Thus, Hypothesis 3b is rejected.

Lastly, to explore the interaction effects of haptic movement and hotspot salience on the usability attributes, a MANCOVA predicting effectiveness, efficiency, and satisfaction as dependent variables (bootstrapped; controlling for age, verbal comprehension, and working memory, see Table 4) was conducted. Results indicated that condition interaction term was a significant predictor of effectiveness ($B = .-1.06$, SE $_{bootstrapped} = .49$) and efficiency ($B = -.12$, SE $_{bootstrapped} = .04$), but not for satisfaction ($B = .-07$, SE $_{bootstrapped} = .19$). Follow-up pairwise comparisons, corrected for Type 1 error, nuanced these patterns (effectiveness: $F(3,124) = 42.49$, p < .00001; efficiency: $F(3,124) = 32.60$, p < .00001). Specifically, children in the salient-hotspot-tapping condition ($M = 9.54$, $SD = .80$) were significantly more effective than all other children. Next in line was the salient-hotspot-dragging condition ($M = 7.97$, $SD = 1.46$) – who were more effective than both non-salient conditions. Children in the non-salient conditions (dragging and tapping) performed the least effectively and similarly to one another (tapping: $M = 6.56$, $SD = 1.69$; dragging: $M = 5.99$, $SD = 1.40$). As to efficiency, children in the salient-drag condition ($M = .85$, $SD = .09$) were significantly more efficient than children in any non-salient condition ($M_{tap} = .56$, $SD = .18$; $M_{drag} = .70$, $SD = .12$). However, they performed similarly to the salient-tap condition ($M = .82$, $SD = .10$). Altogether, evidence provides partial support for Hypothesis 3c.

Discussion

Guided by predictions of the Capacity Model 2.0 and Cognitive Load Theory, this study employed an experimental design to assess the influence of haptic movement (tapping versus dragging; a tactile design feature) and hotspot salience (salient versus non-salient; a visual design feature) on young children's vocabulary learning from an educational app. We hypothesized that the effects of these features would be mediated by usability such that features which were presumed to be less cognitively demanding (i.e. tapping and salient hotspots) would lead to greater usability and improved vocabulary learning. Overall, results indicated that while these design features affected usability, the effects did not translate to learning differences.

No impact of manipulation on learning

This study shows that children were able to learn new words from an educational app in one play session. On average, children learned nearly five words (out of 10) after a single exposure to the app. And this learning occurred regardless of manipulations. Learning was neither directly (nor indirectly via usability) affected by either manipulation. It may be that any changes in learning as a result of design require more repetitive play. In future work, it would be worthwhile to employ a within-subjects design whereby gameplay is repeated at select intervals to identify if and when the effects of design features reveal themselves.

At the same time, games reflect the combination of numerous features, and it may be that any effects would only be detected when features are used in combination. For example, apps which incorporate a series of cognitively-easy features together may produce more robust learning effects, while the use of only one or two such features – although sufficient to move the needle on usability – may not be powerful enough to impact learning. Extending this study in this manner would offer a valuable opportunity to triangulate and extend these findings, particularly when conducted with commercially available content.

Study manipulation considerations

While extending this work is a valuable space for growth, there is another critical space to consider – namely, a critical lens about the study manipulations. Here, we acknowledge feedback from Shalom Fisch regarding our study findings. First, consider the manipulation of tapping versus dragging. We based our predictions on work which showed that while 4-year-olds are able to perform both movements (Nacher et al., 2015), children experience more difficulties with dragging (Aziz, 2013). Russo-Johnson and colleagues had suggested that this may be because dragging is a more complex and requires greater cognitive capacity (Russo-Johnson et al., 2017). Taken together, we hypothesized that tapping would be cognitively easier than dragging. However, this assumption may be wrong. The existing work was based on *computer devices* with *toddlers*. Here, our combination of a *touchscreen* (easier) with *preschoolers* (older) may mean that our manipulation of

cognitive expenditure may have been too small to affect change in learning. In future work, measures which capture expended cognitive resources would be a welcome manipulation check.

At the same time, we also manipulated visual salience via hotspots. Hereto, we take a critical stance. The literature for hotspots suggested that, when connected to educational content, hotspots can be a usable way to direct cognitive resources to the goal content. With this in mind, we expected that salient hotspots would support greater vocabulary learning. However, it could be instead that the salient hotspot prompted a selection heuristic ("pick me!") without encouraging any cognitive encoding. Although usability could (and did) benefit ("this is easy!"), the easiness may have been too easy – resulting in little learning. As one of the reviewers of this manuscript noted, while it is true that animation draws children's attention and can encourage them to easily tap or click, this can obscure any relationship between usability and learning because the same manipulation that makes the present game more usable (namely, the animated hotspot) could prevent children from having to learn the targeted educational content. This is an interesting nuance to the usability-learning relationship that was not sharply considered when designing this study.

In retrospect, an improved manipulation would have been to first provide a moment without haptic feedback to encourage thought – and then provide a salient hotspot as a support cue. Alternatively, and arguably even stronger, one of the manuscript reviewers with experience in educational game design suggested that future work could implement a three-step strategy in design: (1) use hotspots to highlight all response options (correct and incorrect) when the questions are first asked; (2) provide feedback after a first wrong response is made by graying out the option that the child has already selected and highlight the remaining correct and incorrect options, and then (3) providing feedback after a second wrong response by highlighting only the correct answer. Such a change could have a meaningful effect on learning as it benefits usability while also being sensitive to the process of learning. In other words, it is a more appropriate use of the design feature.

Statistical power & individual differences considerations

Lastly, as an additional consideration, we would note that replication of this study with a larger sample size would allow for the detection of statistically small patterns as well as more nuanced analyses that allow deeper examination of individual difference variables – such as working memory capacity (e.g. Choi, Kirkorian, & Pempek, 2021). Indeed, design features may – for some children – benefit learning, while for others that may not (Kirkorian, 2018). It may be that for some children certain design features enhance usability and for other children, they have little, none, or even a negative effect on usability. Even further, for some children, one domain of usability may be more important than another in influencing learning.

Take, for example, our finding for salient hotspots. Salient hotspots were expected to offer a visual aid to help children find the correct answer – which would presumably improve app usability and translate to improved learning. Although children did indicate that the app with salient hotspots was more usable, the step from usability to learning was not detected. In addition to the manipulation caveats noted above, it is also possible

that children with low memory capacity experienced these hotspots as helpful tools that could augment their learning while children with high memory capacity interpreted these hotspots as cues that the game was "easy" and, as a result, allocated less cognitive resources to the game. Early research with television demonstrated that children differentially allocate their resources based on their perceptions of the medium's ease (Salomon, 1984), and there is no reason to think that the same is not true for apps. Even more, when games are experienced as easy, this is likely to decrease engagement (Gee, 2006) which can lead to distraction and less learning – a finding echoed in e-book research (Parish-Morris, Mahajan, Hirsh-Pasek, Golinkoff, & Collins, 2013). Inasmuch, it could be that working memory influenced the effect on learning or even the relationship between usability and learning in such a way that some children benefitted, and others did not – resulting in overall null effects. And as Russo-Johnson and colleagues (Russo-Johnson et al., 2017) found, there are likely other individual differences that might similarly moderate the relationship between design features, usability, and vocabulary learning. Future scholarship which is powered to investigate such fine-grained patterns is valuable next step.

Design features impact usability

This study attempted to spec out one potential route to learning via usability: defined as a multidimensional construct comprised of effectiveness, efficiency, and satisfaction. And while the manipulations themselves were not potent enough to move learning, they began to move the needle on usability – but not always in the same way.

When it comes to effectiveness, as hypothesized, tapping input and salient hotspots best supported usability – particularly when used in combination. Salient hotspots were also found to result in a more efficient app experience. Yet, unexpectedly, tapping resulted in a less efficient app experience. In other words, children required *more* time when tapping as compared to dragging. This pattern contradicts earlier research (Azah & Aziz, 2013; Nacher et al., 2015). However, while tapping was associated with decreased efficiency, the addition of salient hotspots was able to overcome this challenge. In fact, salient hotspots – regardless of tap versus drag combination – predicted the best efficiency. Findings for user satisfaction showed a similar pattern. While neither haptic manipulation influenced satisfaction, children were more satisfied with the app when it relied on salient hotspots. This suggests that hotspot salience may be more influential than haptic input when it comes to efficiency and satisfaction of app play.

While the evidence for hotspot salience is consistent with expectations, the evidence for haptic input was unexpected. Why did tapping take longer? Upon reflection, this might be a consequence of app. In our design, children in the tapping condition had to tap the correct answer twice before moving on to the next screen. The first tap was meant to recognize the correct vegetable, and the second was meant to "put it on the plate". In doing so, this ensured that the child did not provide a correct answer "by accident". However, when children were too quick, the system would interpret it as one tap and expect the child to tap a second time. Although the manipulation was not intended as a *double tap*, children may have interpreted it as such, and young children find the *double*

tap to be very complex to master (Nacher et al., 2015). Consequently, if tapping the screen twice was more complex for children, it would increase the time needed to complete the task thus decreasing the efficiency.

But, considering the fact that tapping did support effective use, this argument for efficiency is not entirely satisfying. Instead, some children may simply need more time to perform the required action (i.e. reduced efficiency), and as a result, experience increased accuracy (i.e. improved effectiveness). Although this seems a logical explanation for our results, it raises questions about the interrelatedness of the different attributes of usability. Based on previous work, we assumed that all three attributes of usability should be supported by cognitively easier features. Yet, this assumption is challengeable. Indeed, there is an ongoing debate as to how different attributes of usability relate to one another (Hornbæk, 2006). Although we did not model usability as one global construct, we did assume similar directionality of relationship between features, usability, and learning. Yet, one could argue that a more realistic conceptualization would suggest that some design features could lead to less efficiency (i.e. more time needed) which, in turn, could predict greater effectiveness – mimicking our haptic input patterns. Increased precision – both in reflecting on the components of usability alongside more precise predictions – would offer important theoretical clarity and practical utility.

Looking ahead

Altogether, this initially simple study is anything but simple. Instead, it is a building-block study to be improved upon in rich ways. It is the first to combine predictions of the CM 2.0 with CLT, and while the expected mediation pathway was not uncovered, this study opens the door for (1) increased theoretical precision, (2) refined methodological practices, and (3) more nuanced analytic decisions. In particular, empirical scholarship which works to better explicate the expected relationships between design features and the three components of usability is well advised. In line with this, lessons here highlight that research with design features requires adaptation sensitivity when going from one medium to another (e.g. computer to touchscreen) as well as measurement of assumed black-box processes. Moreover, replication studies which employ a greater sample size to detect smaller effects and understand boundary conditions are crucial. It is also equally crucial to expand the features investigated as well as consider meta-analyses or systematic reviews to obtain a more comprehensive assessment on the how design features influence learning. And, just as we should expand the features under consideration, more theory and research are necessary to identify which aspects in the process, other than usability, may predict learning outcomes. One can imagine, for example, revised models which not only investigate how design features influence usability, but also consider the role of engagement in the learning process (Ryan & Deci, 2000).

This study also offers important contributions to our present understanding of children's educational apps and future research and development in this space. Most notably, salient hotspots seem to be a particularly potent inroad towards improved usability for young users. This effect seems to be augmented when coupled with other design features (i.e. in this case, tapping) that also require less cognitive resources. From a design perspective, it seems that app designers would be well-advised to incorporate design

features (tactile and visual) which are cognitively easier for their users – relying, for example, on evidence and information from child development experts as well as formative testing. But, as our study shows, usability per se is not a guarantee for learning from educational apps – the route between usability and learning needs to be carefully articulated in the design process. As noted elsewhere, a more thoughtful 3-step salience manipulation may have resulted in more theoretically-consistent findings. We encourage researchers and designers interested in usability to thoughtfully consider *how to* use usability features to support learning. As we saw here, it is not simply whether you have certain features, it is how you use them too.

Lastly, what is usable to one user may be interpreted differently by another. Individual and contextual differences matter, and it will be important in the years to come to identify how and when they influence app effects. In a time and space where technological opportunities are growing exponentially, identifying best practices in children's educational app design will be a moving target. Considering the opportunities of such technology, and its use by even our very youngest children, this is surely a target worth aiming for.

Endnotes

1. Audio features (for example, the presence or type of sound) are also design features – although they are most often used in combination with a visual or tactile feature, as opposed to on their own. In this study, the goal was to focus on one element from the two core design features categories.
2. Apps featured in the work by Chiong and Shuler (2010) were shown to the developer to ensure that the production was stylistically similar.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Jessica Taylor Piotrowski is an Associate Professor in the Amsterdam School of Communication Research at the University of Amsterdam (UvA). She directs the Center for research on Children, Adolescents, and the Media (CcaM) where she is presently studying digital competence. She is the co-author of the book *Plugged in: how media attract and affect youth* (Yale University Press, 2017), and regularly publishes in communication, psychology, and education journals.

Francette L. Broekman is a Research & Management Assistant at Leiden University. She is specialized in e-learning with a focus on youth, education, and entertainment. This project formed part of her doctoral thesis, entitled "Tap in, swipe through, pinch out: understanding the role of context, content, and child in apps for children".

ORCID

Jessica Taylor Piotrowski 🆔 http://orcid.org/0000-0002-2756-5197

References

Aziz, N. A. A. (2013). Children's interaction with tablet applications: Gestures and interface design. *Children*, *2*(3), 447–450.

Bedford, R., Saez de Urabain, I. R., Cheung, C. H. M., Karmiloff-Smith, A., & Smith, T. J. (2016). Toddlers' fine motor milestone achievement is associated with early touchscreen scrolling. *Frontiers in Psychology*, *7*(1108). doi:10.3389/fpsyg.2016.01108

Bevan, N. (2006). Practical issues in usability measurement. *Interactions*, *13*(6), 42–43.

Bevan, N., Carter, J., & Harker, S. (2015). ISO 9241-11 revised: What have we learnt about usability since 1998?. In International conference on human-computer interaction (pp. 143–151). Springer, Cham.

Brooke, J. (2013). SUS: a retrospective. *Journal of usability studies*, *8*(2), 29–40.

Chiong, C., & DeLoache, J. S. (2013). Learning the ABCs: What kinds of picture books facilitate young children's learning? *Journal of Early Childhood Literacy*, *13*(2), 225–241.

Chiong, C., & Shuler, C. (2010). *Learning: Is there an app for that? Investigations of young children's usage and learning with mobile devices and apps* Retrieved September 2018, from http://www-tc.pbskids.org/read/files/cooney_learning_apps.pdf

Choi, K., Kirkorian, H. L., & Pempek, T. A. (2021). Touchscreens for whom? Working memory and age moderate the impact of contingency on Toddlers' transfer from video. *Frontiers in Psychology*, 12. doi:10.3389/fpsyg.2021.621372

Common Sense Media. (2017). *The common sense census: Media use by kids age zero to eight 2017*. Retrieved from San Francisco, CA: Common Sense Media.

de Jong, M. T., & Bus, A. G. (2002). Quality of book-reading matters for emergent readers: An experiment with the same book in a regular or electronic format. *Journal of Educational Psychology*, *94*(1), 145–155.

De Jong, M. T., & Bus, A. G. (2003). How well suited are electronic books to supporting literacy? *Journal of Early Childhood Literacy*, *3*(2), 147–164.

de Jong, M. T., & Bus, A. G. (2004). The efficacy of electronic books in fostering kindergarten children's emergent story understanding. *Reading Research Quarterly*, *39*(4), 378–393.

Fisch, S. M. (2000). A capacity model of children's comprehension of educational content on television. *Media Psychology*, *2*, 63–91.

Fisch, S. M. (2004). *Children's learning from educational television*. Mahwah, NJ: Erlbaum.

Fisch, S. M. (2016). The capacity model 2.0: Cognitive processing in children's comprehension of educational games. In *Paper presented at the Society for Research in Children's Development*. Irvine, CA: Special Topics Media on Technology and Children's Development.

Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000). *Measuring usability: Are effectiveness, efficiency, and satisfaction really correlated?* Paper presented at the Proceedings of the SIGCHI conference on Human Factors in Computing Systems The Hague, The Netherlands.

Gagne, R. M., & Briggs, L. J. (1974). *Principles of instructional design*. New York: Holt, Rinehart & Winston.

Gee, J. P. (2006). Learning by design: Good video games as learning machines Messaris, Paul, Humphreys, Lee. In *Digital media: Transformations in human communication* (pp. 173–186). New York: Peter Lang.

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, *76*(4), 408–420.

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.

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

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*, *64*(2), 79–102.

Kiili, K., De Freitas, S., Arnab, S., & Lainema, T. (2012). The design principles for flow experience in educational games. *Procedia Computer Science*, *15*, 78–91.

Kirkorian, H. L. (2018). When and how do interactive digital media help children connect what they see on and off the screen? *Child Development Perspectives*, *12*(3), 210–214.

Kirkorian, H. L., Choi, K., & Pempek, T. A. (2016). Toddlers' word learning from contingent and non-contingent video on touchscreens. *Child Development*, *87*(2), 405–413.

Korat, O., & Shamir, A. (2007). Electronic books versus adult readers: Effects on children's emergent literacy as a function of social class. *Journal of Computer Assisted Learning*, *23*(3), 248–259.

Laurillard, D. (2009). The pedagogical challenges to collaborative technologies. *International Journal of Computer-Supported Collaborative Learning*, *4*(1), 5–20.

Lin, L.-Y., Cherng, R.-J., & Chen, Y.-J. (2017). Effect of touch screen tablet use on fine motor development of young children. *Physical & Occupational Therapy in Pediatrics*, *37*(5), 457–467.

Linebarger, D. L., Moses, A., & McMenamin, K. (2010). *Vocabulary learning from educational television: Can children learn new words from Martha speaks?* Philadelphia, PA: Annenberg School for Communication.

Linebarger, D. L., Piotrowski, J. T., & Vaala, S. (2007). *Vocabulary knowledge, conversational language skills, and map knowledge: Can Postcards from Buster help inside the ELL classroom?* Philadelphia, PA: .Annenberg School for Communication.

Marsh, J., Plowman, L., Yamada-Rice, D., Bishop, J., Lahmar, J., Scott, F., . . . Piras, M. (2015). *Exploring play and creativity in pre-schoolers' use of apps: Final project report*. Retrieved from

Merrill, M. D., Drake, L., Lacy, M. J., Pratt, J., & Group, I. R. (1996). Reclaiming instructional design. *Educational Technology*, *36*(5), 5–7.

Nacher, V., Jaen, J., Navarro, E., Catala, A., & González, P. (2015). Multi-touch gestures for pre-kindergarten children. *International Journal of Human-Computer Studies*, *73*, 37–51.

Neumann, M. M. (2014). An examination of touch screen tablets and emergent literacy in Australian pre-school children. *Australian Journal of Education*, *58*(2), 109–122.

Neumann, M. M. (2018). Using tablets and apps to enhance emergent literacy skills in young children. *Early Childhood Research Quarterly*, *42*, 239–246.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*(1), 63–71.

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*(3), 200–211.

Piotrowski, J. T., & Krcmar, M. (2017). Reading with hotspots: Young children's responses to touchscreen stories. *Computers in Human Behavior*, *70*, 328–334.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891.

Reigeluth, C. M., & Carr-Chellman, A. A. (2009). *Instructional-design theories and models, volume III: Building a common knowledge base*. UK: Routledge.

Roman, A. S., Pisoni, D. B., & Kronenberger, W. G. (2014). Assessment of working memory capacity in preschool children using the missing scan task. *Infant and child development*, *23*(6), 575–587

Ronimus, M., Kujala, J., Tolvanen, A., & Lyytinen, H. (2014). Children's engagement during digital game-based learning of reading: The effects of time, rewards, and challenge. *Computers & Education*, *71*, 237–246.

Roskos, K., Brueck, J., & Widman, S. (2009). Investigating analytic tools for e-book design in early literacy learning. *Journal of Interactive Online Learning*, (3), 218–240.

Russo-Johnson, C., Troseth, G., Duncan, C., & Mesghina, A. (2017). All tapped out: Touchscreen interactivity and young children's word learning. *Frontiers in Psychology*, *8*, 1–15.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78.

Salomon, G. (1984). Television is "easy" and print is "tough": The differential investment of mental effort as a function of perceptions and attributions. *Journal of Educational Psychology*, *76*, 647–658.

Sauro, J., & Kindlund, E. (2005). How long should a task take? identifying specification limits for task times in usability tests. In Proceeding of the Human Computer Interaction International Conference (HCII 2005), Las Vegas, USA.

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, *19*(4), 469–508.

Shamir, A. (2009). Processes and outcomes of joint activity with e-books for promoting kindergarteners' emergent literacy. *Educational Media International*, *46*(1), 81–96.

Smeets, D. J., Van Dijken, M. J., & Bus, A. G. (2014). Using electronic storybooks to support word learning in children with severe language impairments. *Journal of Learning Disabilities*, *47*(5), 435–449.

Statistica. (2019). Number of apps available in leading app stores as of 1st quarter 2019 Retrieved September 2019, from https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/

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.

Vaala, S., Ly, A., & Levine, M. H. (2015). *Getting a read on the app stores: A market scan and analysis of children's literacy apps*. Retrieved from New York City, NY, USA: The Joan Ganz Cooney Center at Sesame Workshop.

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development*, *69*(3), 848–872.