Computers in Human Behavior 59 (2016) 58–66

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

Achieving the performance benefits of hands-on experience when using digital devices: A representational approach

Sara G. Goodman*, Travis L. Seymour, Barrett R. Anderson

University of California Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA

ARTICLE INFO

Article history:
Received 29 October 2015
Received in revised form
4 January 2016
Accepted 6 January 2016
Available online xxx

Keywords:
Mobile learning
Human computing interaction
Mathematics
Tablet computers
Representation

ABSTRACT

Digital devices have become ubiquitous fixtures in classrooms nationwide. Despite this rapid incorporation of tablet computers in educational settings, the costs and benefits of digitization are understudied. Prior research shows that reading comprehension tends to be best in physical rather than digital modalities. The current study extends these findings to physical and digital versions of spatial tasks. Participants engaged in a physical or digital tangram task and demonstrated significant performance differences in both accuracy and response time. Later, on a timed math test, participants solved more problems correctly and solved each problem more quickly if physical manipulatives, rather than a digital interface, were used in the tangram task. Two follow-up experiments showed that these performance differences are not due to interactional limitations of the tablet, but are instead likely driven by representational differences. These results show that priming physical representations of spatial tasks can reduce the digital task performance deficit.

1. Introduction

More than 1/3 of US middle and high school students now use school issued mobile devices (e.g., smartphones and tablets), and 60–70% do schoolwork on some type of mobile device (Speak Up 2013 survey; Project Tomorrow, 2014). The rise of this technology in the classroom has led to an interest in how performance and engagement compares between more traditional interfaces and newer digital variations (Esteves, van den Hoven, & Oakley, 2013; Manches, O'Malley, & Benford, 2009). Digital variants of traditional learning materials can offer many unique and useful features. For example, digital books can include audio, video and web-links that allow for potentially richer learning experiences than print books. Other benefits include enhanced cross-reference and search abilities, increased portability, consistency of interface across multiple books, cheaper revisions and more efficient distribution. However, digital books and other digital learning materials may also lack some features that enhance learning compared to traditional analog versions.

One important difference between desktop and mobile computer based learning tools and previous incarnations is in the level of physical interaction and manipulation involved. Whether a book or other learning tool is physical or digital, it becomes integrated as a part of one's representation and understanding of the educational task. The representation of the task is built upon the individual's internal representation of the task space, the external manifestation of the task itself (e.g. physically manipulable task, digital interface), and existing knowledge of relevant factors. According to theories of *distributed cognition*, the individual must assemble these separate internal and external representations, memory for task rules, and other key components into a cohesive problem space that can be used to complete the task (Zhang, 1997; Zhang & Norman, 1994). The nature of both the internal and external representations can greatly influence one's strategy, exploration, and understanding of the task's constraints (Kirsh & Maglio, 1994; Zhang, 1997; Zhang & Norman, 1994). As a direct result, an individual's subsequent performance on the task is affected by the ability to construct an accurate and enriched representation of the problem. Changes in the nature of the task can influence the problem representation, and thereby influence performance. For example, moving from the physical version of a task where one manipulates 3 dimensional objects with one's hands to an online version where one manipulates 2 dimensional representations indirectly via a mouse (or finger gestures) may represent a significant enough change in interface and external

* Corresponding author. Psychology Department, University of California Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA.
E-mail address: slgreen@ucsc.edu (S.G. Goodman).

http://dx.doi.org/10.1016/j.chb.2016.01.006
0747-5632/

representation that task performance is affected. Although theories regarding distributed cognition do not predict whether this will lead to better or worse performance, these theories do suggest that performance will ultimately be affected.

Following this theoretical formulation, a growing body of research has examined the effect of performing various tasks when they are presented with interactive and tangible user interfaces (TUIs) compared to graphical user interfaces (GUIs) with minimal interactive capabilities and indirect object manipulation via computer mouse. This type of head-to-head comparison more often shows performance advantages for TUIs and tasks with more tangible components, whereas performance is shown to be worse for GUIs and less interactive tasks (Do-Lenh, Jermann, Cuendet, Zufferey, & Dillenbourg, 2010; Fitzmaurice & Buxton, 1997; Jacob, Ishii, Pangaro, & Patten, 2002; Manches et al., 2009; Patten & Ishii, 2000; Schneider, Jermann, Zufferey, & Dillenbourg, 2011; Terrenghi, Kirk, Sellen, & Izadi, 2007; Tuddenham, Kirk, & Izadi, 2010; Xie, Antle, & Motamedi, 2008; see Zuckerman & Gal-Oz, 2013; for review). In studies that do not report differences across interfaces, there are still observable differences in terms of strategic planning, exploration-directed or epistemic actions, and problem solving efficiency (Esteves et al., 2013; Marshall, Cheng, & Luckin, 2010; Stull, Barrett, & Hegarty, 2013). For example, in a recent study by Esteves et al. (2013), participants played a four-in-a-row matching game, using a mouse-based interface, a touch-based interface, and a tangible (physical token) based interface. No significant differences were reported in task performance between interfaces, however, there were differences in both the number of moves ahead that participants reported considering, and in the number and types of epistemic actions taken between interfaces. Similarly, a study directly comparing performance using virtual and concrete models in chemistry instruction (Stull et al., 2013) found an improvement in efficiency for participants using the virtual models, which facilitated subsequent efficiency when using concrete models. Likewise, a study of spatial planning using interfaces with different degrees of physicality found an interface-dependent difference in the number of epistemic actions executed by the participant (Fjeld & Barendregt, 2009). These differences may be because physically embodied interactive elements allow for a greater variety of trial-and-error exploration. In contrast, virtual interfaces tend to have simpler affordances that might guide users to specific task-relevant actions. In all, these studies generally support the distributed cognition account, whereby differences in the qualities of a given external representation are integrated into separate and qualitatively different problem spaces.

In congruence with the findings from the TUI research, educators have employed physical manipulatives to teach a variety of topics since the 19th century (Brosterman, 1997). These manipulatives provide concrete examples of abstract concepts and are used as an alternative to relying on abstract symbolic systems alone. For example, children at early educational stages might learn about counting by moving a collection of beads one or two at a time, or develop an understanding of comparison by moving weights on or off of a scale. In later educational stages, play money is often used to facilitate learning about addition and subtraction, whereas experimentation with a set of blocks cut to different sizes can be used to teach fractions. Although these methods have been subject to skepticism (McNeil & Jarvin, 2007; Moyer, 2001), a recent meta-analysis found that the use of physical manipulatives in math education tends to improve retention, problem solving, and transfer (Carbonneau, Marley, & Selig, 2013).

In related technology-based educational trends, the growing use of computer and mobile devices is moving instruction away from tangible interfaces and towards screen based ones. One of the most prevalent changes is the growing use of e-book and other digital reading materials rather than printed books. This shift toward digital reading material has motivated several studies that examine the effects of this change on students' reading comprehension. These studies have shown that in contrast to digital text, reading physical books leads to improved comprehension (Ackerman & Goldsmith, 2011; Ackerman & Lauterman, 2012; Chen, Cheng, Chang, Zheng, & Huang, 2014). Ackerman and Goldsmith (2011) showed that college-aged participants who read long passages on a computer screen scored more poorly on a comprehension test than participants who read the passages on paper when given a time constraint. Yet, these participants reported higher levels of confidence regarding their understanding of the material. Similar results were obtained in a follow-up study that manipulated reading format (screen vs. paper) using a within-subjects design (Ackerman & Lauterman, 2012). A similar study showed that 10th grade students also scored higher on reading comprehension for lengthy texts that had been presented on paper than on-screen (Mangen, Walgermo, & Brønnick, 2013). In addition to these comprehension differences, printed material yielded advantages for engagement with the text as well as an understanding of the temporal ordering of events when compared to digital material (Mangen & Kuiken, 2014). Not surprisingly, physical books are also reported as the overwhelming favorite modality among readers (Buzzetto-More, Sweat-Guy, & Elobaid, 2007; Spencer, 2006; Woody, Daniel, & Baker, 2010).

A similar trend is emerging in education with an increase in the prevalence of digital representations on tablets and computers used for teaching spatial skills, in contrast with the more traditional physical manipulatives. For example, the Tangram puzzle task (Fig. 1) involves presenting participants with a standard set of shapes and a target pattern. The goal is to rotate, flip and arrange the individual pieces to create the pattern. In addition to generally improving spatial skills, several studies have shown a direct link between such spatial skills and math performance (Burnett, Lane, & Dratt, 1979; Casey, Nuttall, & Pezaris, 2001; Delgado & Prieto, 2004; Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Geary, Saults, Liu, & Hoard, 2000; Holmes, Adams, & Hamilton, 2008; Kyttälä, Aunio, Lehto, Van Luit, & Hautamäki, 2003; McKenzie, Bull, & Gray, 2003; McLean & Hitch, 1999; Rasmussen & Bisanz, 2005; Thompson, Nuerk, Moeller, & Kadosh, 2013). For example, Cheng and Mix (2014) showed that for 6–8 year olds, practicing mental rotation led to improved performance on missing term ($2+\_ = 11?$) problems (for a contrary view, see Hawes, Moss, Caswell, & Poliszczuk, 2015). Cheng and Mix argued that the observed relationship between mental rotation and math performance stems from students' attempts to mentally rotate missing term problems (e.g. $4 + \_\_ = 12$) into more familiar layouts (e.g. $12 - 4 = \_\_$). Others have also demonstrated that spatial training, including free play with blocks or physically enacting motion along a number line mat, can improve math performance (Fischer, Moeller, Bientzle, Cress, & Nuerk, 2011; Wolfgang, Stannard, & Jones, 2001).

As with reading, teaching spatial skills using computerized tangram puzzles may significantly alter one's internal and external representations of the task and its constraints, as well as affecting the way participants approach the problem space. Thus, tangram performance may be influenced by the use of a digital interface when compared to a physical one. Perhaps more importantly, it is also unknown whether digital tangram interfaces will magnify or attenuate the degree to which the resulting spatial practice will enhance math performance as is reported following physical tangram practice.

The present study was designed to revisit the question of whether a shift from paper to screens has a negative impact on performance, but is the first to ask this question with a spatial problem solving task. This study is also the first to examine

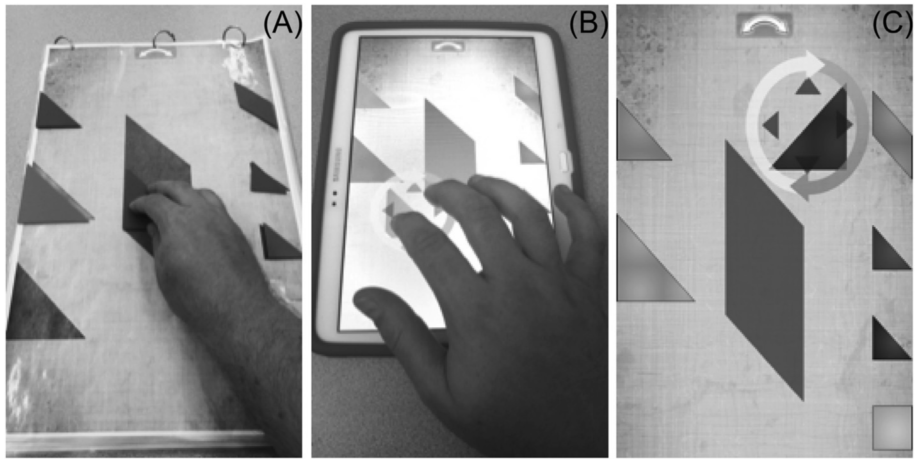

Fig. 1. (A) Example starting puzzle setup for the tabletop task. (B) Example starting puzzle setup for the tablet task. (C) Tablet user interface for manipulating virtual tangram shape tiles.

whether such a change from tangible to digital interfaces may additionally attenuate the degree to which the task has knock-on educational benefits typically reported for the tangible version.

Finally, in nearly all previous studies that show better performance for tangible rather than digital interfaces, continued use of tangible interfaces is clearly suggested. This can be achieved either by using the traditional hands-on version, creating a digital version with an interface that mimics the critical aspects of the physical one, or modifying the task in such a way as to sufficiently enrich one's task representation. If hands-on tangram puzzle practice leads to better performance than a digital analogue, the present study is the first to examine whether it is possible to make simple changes to the digital task that will mitigate the performance decrement and/or the knock-on math effect that would otherwise be missing in the digital version.

1.1. Analysis plan

To test stated hypotheses in the following experiments, two-tailed t-tests were conducted on each planned contrast with a 0.05 alpha level for significance. For each contrast, the 95% confidence interval and effect size (Cohen's D) is also reported. Rather than remove outlying data, distributions of puzzle solution times were winsorized using a 10% trim (Erceg-Hurn & Mirosevich, 2008). For each of the three between-subjects experiments, an a priori power analysis using the G*Power software (version 3.1) indicated a need for 88 subjects in each of the two groups in order to reach 95% power for detecting a medium effect at a significance level of 0.05 (Faul, Erdfelder, Lang, & Buchner, 2007).

2. Experiment 1

Experiment 1 assessed performance on a tangram puzzle task for participants engaging in a hands-on physical version or a digitized tablet computer version. Based on previous research showing the advantages of tangible interfaces in general (e.g., Do-Lenh et al., 2010; Fitzmaurice & Buxton, 1997; Jacob et al., 2002; Manches et al., 2009; Patten & Ishii, 2000; Schneider et al., 2011; Terrengi et al., 2007; Tuddenham et al., 2010; Xie et al., 2008), and also related studies examining the benefits of paper-based reading over screen-based reading (e.g. Ackerman & Goldsmith, 2011; Ackerman & Lauterman, 2012; Chen et al., 2014), we predicted that performance on the physical tangram task would be better than digital tangram performance. Experiment 1 also attempted to replicate previous results showing increased math performance after spatial skill training (e.g., Cheng & Mix, 2014; Fischer et al., 2011; Wolfgang et al., 2001) and also examined whether this benefit would still hold when the preceding spatial task was digital. We predicted that tangram practice would lead to a greater benefit on subsequent math performance in the physical than digital condition.

2.1. Methods

2.1.1. Participants

One hundred eighty undergraduates (126 female, 54 male) were recruited for this study from the University of California Santa Cruz and received course credit for their participation. Participant ages ranged in from 17 to 29 (*M*: 19.49, *SD*: 1.62).

2.1.2. Materials

Tablet-based tangram puzzles were completed on a Samsung Galaxy Tab 3 10.1 inch Android OS based tablet computer using the Tangram HD application version 3.0 created by Pocket Storm. For the tabletop tangram task, the target shape and allowed shapes were digitized, enlarged, and printed from the tablet application. Tiles for this task were plastic tangram shapes produced by Learning Resources, Inc. Shape tiles consisted of 2 large triangles (7 cm by 7 cm), 1 medium triangle (5 cm by 5 cm), 2 small triangles (3.5 cm by 3.5 cm), 1 square (3.5 cm by 3.5 cm), and 1 parallelogram (5 cm by 2.5 cm). Timing was done using an Accusplit Pro Survivor 601X digital stopwatch.

2.1.3. Procedure

Participants were asked to complete a series of tangram puzzles by arranging, rotating and flipping a set of seven small geometric shape tiles to match a larger composite shape. For example, Fig. 1 shows the template for a vertical parallelogram and the peripheral starting position for each shape tile. Participants were randomly assigned to one of two conditions. Half were assigned to a tangible tabletop condition that featured the use of plastic versions of the geometric manipulatives. In this condition, shape tile manipulation involved using the hands in stereotypical sliding, rotating, and flipping motions (see Fig. 1a). The remaining participants were assigned to the tablet-based condition, in which the same puzzles were completed via a tangram application on a 10 inch tablet-computer screen (Fig. 1b). Here, shape tile manipulation involved an interaction style typical for tablet applications, but unusual for table-top puzzles. To select a tile for interaction, one touched the tile, which revealed an interface (Fig. 1c) that allowed moving (dragging the virtual tile with one's index finger across the display screen) and rotating (circling one's index finger around the virtual tile until it had been rotated the desired amount). Finally, flipping a tile involved pressing the dedicated button located in the top-center of the display while a shape tile was highlighted.

Prior to the start of the first puzzle, a researcher slowly demonstrated the correct solution for an example puzzle. Afterwards, this puzzle was reset and participants were asked to recreate the solution. After completing the demonstration puzzle, participants were instructed to solve each tangram as quickly as possible. Researchers used digital stopwatches to record the solution time for each tangram. Participants continued the tangram task until all 20 puzzles had been solved or 40 min had elapsed. Immediately following the tangram task, participants were asked to complete a timed math test using paper and a pencil that consisted of 25 addition and 25 subtraction problems. In a similar approach as the one used by Cheng and Mix (2014), who suggested that the benefit of spatial practice on math performance may be linked to mentally rotating missing term problems to a familiar format, half of each type of question was depicted horizontally (e.g., 384 + 146__) and half were depicted vertically (e.g., $\begin{array}{r}384\\+146\\\hline\end{array}$), requiring that participants engage in mental rotation to orient the problems in the preferred manner. Researchers used digital stopwatches to measure math test completion time.

2.2. Results

As predicted, participants in the Tablet condition solved fewer puzzles, $t(171) = -3.06$, $p = 0.003$, $d = 0.45$, and solved each puzzle more slowly, $t(178) = 3.88$, $p < 0.001$, $d = 0.58$, than those in the Tabletop condition. Performance on the subsequent math test was also consistent with predictions; participants in the Tablet condition correctly answered fewer problems, $t(118) = -3.15$, $p = 0.002$, $d = 0.48$, and finished the math test more slowly, $t(178) = 3.49$, $p = 0.001$, $d = 0.52$, than those in the Tabletop condition. See Table 1 for means, standard deviations, and confidence intervals for tested contrasts.

2.3. Discussion

The performance differences observed in this experiment are consistent with previous work showing an advantage for tangible interfaces over more indirect ones (e.g., mouse-based; see Zuckerman & Gal-Oz, 2013 for review). However, in addition to the difference in tangibility between conditions, the tabletop and tablet tasks also differed significantly in the afforded interactional style. In particular, the tabletop task offered participants a well-practiced and familiar interaction style involving hand manipulations humans use for a variety of tasks. However, the tablet-interface restricted participants to moving, turning, and flipping the virtual objects using single finger gestures. Furthermore, flipping a shape tile required selecting it and pressing a separate flip button. Because this function only flips the piece around the vertical axis, horizontal flipping also required a rotation. Thus, it is possible that significant differences in task interface, rather than tangibility alone, may be driving the puzzle solution differences, and potentially even the subsequent differences in math performance.

3. Experiment 2

In Experiment 2, we tested the idea that the unfamiliar and unconventional manner in which shape tiles are manipulated in the tablet-based tangram task, and not differences in tangibility per se, led to the tangram and math performance results observed in Experiment 1. This was achieved by replicating Experiment 1's tabletop condition, but asking half of participants to adopt a restricted 2-finger interaction style reminiscent of the style participants used on the tablet interface (Fig. 2 right panel). The other half of participants used the normal unrestricted tabletop tangram manipulation style (stereotypical full-hand rotation, moving, and flipping, see Fig. 2 left panel). If the tabletop interface drove Experiment 1 results, we should see greater performance from those using the normal tabletop interaction style compared to those limited to the tablet interaction style. We predicted that the constraint on interaction style would result in a comparable performance difference to the one observed in Experiment 1, suggesting that those findings are due to the limited (or unfamiliar) user interface in the tablet condition relative to the tabletop condition.

3.1. Methods

3.1.1. Participants

One hundred thirty seven undergraduates (101 female, 36 male)

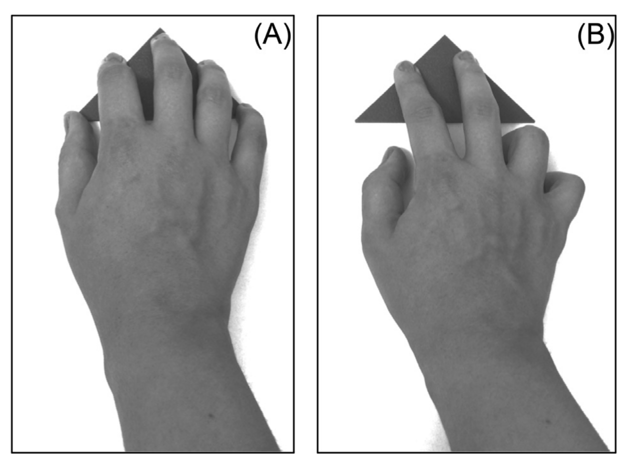

Fig. 2. (A) Unrestricted (standard) shape–tile interaction style. (B) Restricted 2-finger interaction style.

Table 1
Mean performance scores for tangram and math test tasks across Tablet and Tabletop conditions.

	Tablet condition		Tabletop condition		95% CI (Difference)
	M	*SD*	*M*	*SD*	
Puzzles solved	8.28	3.96	10.38	5.18	[−3.44, −0.74]*
Puzzle solution time (sec)	179.03	63.91	140.08	70.72	[19.15, 58.76]*
Math number correct	45.64	4.83	47.42	2.17	[−2.89, −0.66]*
Math completion time (sec)	466.29	133.02	394.83	141.97	[31.02, 111.91]*

Note. CI = confidence interval.
* $p < 0.05$.

were recruited for this study from the University of California Santa Cruz and received course credit for their participation. Participant ages ranged from 17 to 26 (*M*: 19.78, *SD*: 1.84).

3.1.2. Materials

The materials for this experiment were the same as those described for the tabletop condition in Experiment 1.

3.1.3. Procedure

Participants were randomly assigned to either the Restricted or Unrestricted shape tile interaction style. Prior to the start of the first puzzle, during the demonstrated solution for an example puzzle, participants were asked to also use the demonstrated interaction style. If unrestricted tactics were observed for participants in the Restricted condition, they were immediately asked to resume using the restricted style. Otherwise, the method and procedure for Experiment 2 was identical to that described for the tabletop condition in Experiment 1.

3.2. Results

Contrary to expectations, the restricted interaction afforded by the tablet interface does not appear to have contributed to the results of Experiment 1. In Experiment 2, participants in the Restricted interaction condition solved the same number of puzzles, $t(134) = 0.36$, $p = 0.722$, $d = 0.06$, and solved them with equal speed, $t(133) = -0.37$, $p = 0.71$, $d = 0.06$, as those in the Unrestricted interaction condition. Performance on the subsequent math test was also identical across conditions. Participants in the Restricted condition correctly answered the same number of problems, $t(135) = -0.01$, $p = 0.99$, $d = 0$, and finished the math test with equal speed, $t(131) = 1.53$, $p = 0.13$, $d = 0.26$, compared to those in the Unrestricted condition. See Table 2 for means, standard deviations, and confidence intervals for tested contrasts.

3.3. Discussion

The results from Experiment 2 are based on a lack of difference in the four contrasts, and arguing from null results should be done with care. However, it is important to recall that the sample size was based on a previous power analysis, and that a nearly identical paradigm and measures were used in Experiment 1. Furthermore, the confidence intervals for each contrast cross 0, which indicates that there is no clear pattern in the data to support a consistent benefit or detriment for the restricted vs. unrestricted interaction style.

Based on the present results, the puzzle solving performance decrement observed when participants solve tangrams in the tablet condition (compared to Tabletop) is not likely to have been caused by the physical interactional constraints of the tablet. Previous research has shown that problem solving strategies differ as a result of the affordances that are primed across virtual and physical objects (Fjeld & Barendregt, 2009; Manches et al., 2009). Further, the affordances of a given interface sway the user toward engagement in different types of information-seeking actions, and can consequently influence the developed representations of the task over time (Esteves et al., 2013; Marshall et al., 2010; Stull et al., 2013). Thus, the driving factor for the performance differences between tangible (e.g., using physical tangram shape tiles) and indirect (e.g., using virtual tangram shape tiles) interfaces may have been due to induced differences in participants overall representation for the task. Presumably, this representation difference also led to the observed effect on subsequent math performance. To test this conclusion, we would need to manipulate participants' internal representation of the task components and examine its influence on tangram performance.

4. Experiment 3

Experiment 3 was designed to test whether the performance differences across interfaces observed in Experiment 1 were due to differences in participants' internal representation of the task and in particular, the shape tiles. Using the tablet version of the paradigm, we manipulated whether or not each tangram tile's, shape, size, and 3D rotation were primed prior to the tangram task. Results from Experiments 1 and 2 suggest that observed performance differences between tabletop and tablet conditions stemmed from differences in internal task-representation. Thus, we expected that tangram performance, and subsequent math performance, would increase as the shape and manipulability of the shape tiles were primed.

4.1. Methods

4.1.1. Participants

Participants were 217 undergraduates (161 female, 56 male) who were recruited from the University of California Santa Cruz. All participants received course credit for their participation. Participant ages ranged in from 18 to 32 (M: 19.54, SD: 1.71). (task-relevant or task-irrelevant) prior to completing tangrams using the digital interface.

4.1.2. Materials

The materials for this experiment are identical to those described for the Tablet condition in Experiment 1, except for the addition of two new video presentations. A tangram-characteristic video provided a visual demonstration of the characteristics of each tangram piece along with an auditory explanation. For each piece, the video described the number of sides, length of sides ("short", "long") and number of angles by type ("acute", "right", "obtuse"). The demonstrator featured in the video uses his hands to rotate each piece to the left and to the right, and also to invert each piece. A leaf-characteristic video used the same explanatory structure as the tangram video, but described the characteristics of five different tree leaves. The leaf video described the symmetry and edge characteristics of each leaf and intentionally contained no depiction

Table 2
Mean performance scores for tangram and math test tasks across Restricted and Unrestricted rotation conditions.

	Restricted		Unrestricted		95% CI (Difference)
	M	*SD*	*M*	*SD*	
Puzzles solved	7.81	3.39	7.61	3.18	[−0.91, 1.31]
Puzzle solution time (sec)	163.46	75.1	167.97	66.39	[−28.48, 19.46]
Math number correct	46.01	3.88	46.02	4.14	[−1.36, 1.35]
Math completion time (sec)	461.09	108.05	429.69	131.82	[−9.32, 72.11]

Note. CI = confidence interval.

of hands or implied how one might use their hands to examine the leaves. Each video was shown on a Hewlett Packard L1740 17 inch (37.39 cm × 38.1 cm) LCD color monitor (1280 × 1024 pixels). Each video was presented in full-screen mode through a web-based video presentation system. The Unprimed condition video lasted for 2 min and 29 s. The Priming condition video lasted for 3 min and 42 s.

4.1.3. Procedure

Participants were randomly assigned to one of two priming conditions: half were assigned to an Unprimed condition in which they watched the leaf-characteristic video (Fig. 3a) and half were assigned to a Task-Relevant priming condition, in which they watched the tangram-characteristic instructional video about the shape tiles prior to engaging in the tangram task (Fig. 3b). Following the video prime, the procedure for Experiment 3 was identical to the tablet tangram condition in Experiment 1.

4.2. Results

As predicted, participants in the Unprimed representation condition solved fewer puzzles, $t(162) = -2.56$, $p = 0.012$, $d = 0.36$, and took more time on average to solve each puzzle, $t(139) = 2.22$, $p = 0.028$, $d = 0.33$, compared to those in the Primed representation condition. Contrary to prediction, participants in the Unprimed representation condition did not successfully answer fewer problems on the subsequent math test, $t(113) = -1.82$, $p = 0.072$, $d = 0.29$. However, consistent with predictions, participants in the Unprimed condition did answer math questions more slowly than those in the Primed condition, $t(149) = 2.14$, $p = 0.034$, $d = 0.31$. See Table 3 for means, standard deviations, and confidence intervals for tested contrasts.

4.3. Discussion

Results from Experiment 3 clearly demonstrate that enhanced representation of the individual geometric puzzle pieces and associated motor actions, achieved through visual priming, significantly improves performance on the digital task. Thus, by merely presenting participants with a pre-task instructional video, the representations of the pieces and motor actions that contribute to the problem space for the digital task can be enhanced artificially and performance can be improved. In accordance with the previously discussed theories of distributed cognition (Kirsh & Maglio, 1994; Zhang, 1997; Zhang & Norman, 1994), this experiment demonstrates that the digital form of the tangram task lacks sufficient representational activation of the physical characteristics of manipulatives, as well as activation of the ways in which one may physically interact with the pieces. Thus, these results also provide a clear suggestion for a simple and effective way to alleviate the digital interface performance deficit.

5. Post-Hoc comparison

These three experiments present compelling evidence for the role of representation in performance differences across table-top and tablet interfaces. However, it is important to note that the table-top setup is larger than the tablet interface, and this difference in size must be considered as a potential factor contributing to this performance difference. Regardless of the role of interface size, a comparison of the Unprimed tablet condition from Experiment 3 and the Table-top condition from Experiment 1 should reveal the same pattern of performance differences originally observed in Experiment 1, thus equating the Unprimed tablet condition in Experiment 3 with the Tablet condition in Experiment 1. However, if the size of the interface rather than one's mental representation of the task is responsible for these performance differences, the comparison between the Primed tablet condition from Experiment 3 and the Table-top condition from Experiment 1 should indicate some degree of performance difference, suggesting that the priming of a representation is not sufficient to alleviate the performance differences between tablet and table-top interfaces.

As predicted, participants in the Unprimed tablet condition (Experiment 3) solved marginally fewer puzzles, $t(161) = -1.85$, $p = 0.066$, $d = 0.28$, and took more time on average to solve each puzzle, $t(160) = 2.06$, $p = 0.041$, $d = 0.31$, compared to those in the Table-top condition (Experiment 1). Participants in the Unprimed tablet condition did not answer fewer problems on the subsequent math test, $t(141) = -0.89$, $p = 0.377$, $d = 0.14$. However, consistent with predictions, participants in the Unprimed tablet condition did answer math questions more slowly than those in the Table-top condition, $t(166) = 3.28$, $p = 0.001$, $d = 0.5$. See Table 4 for means, standard deviations, and confidence intervals for tested contrasts.

Contrary to the expectation that display size accounts for some of the difference in task performance, there is no difference in tangram or math problem performance between participants in the Primed tablet condition from Experiment 3 and participants in the Table-top condition from Experiment 1. Participants in these conditions solved the same number of puzzles, $t(155) = 0.11$, $p = 0.912$, $d = 0.02$, and did not differ in the average time it took to solve each puzzle, $t(135) = 0.54$, $p = 0.587$, $d = 0.08$. Further, participants did not differ in the number of math problems they solved, $t(176) = 1.07$, $p = 0.288$, $d = 0.15$, or the average time it took to complete the math test, $t(168) = 1/69$, $p = 0.092$, $d = 0.24$. See Table 5 for means, standard deviations, and confidence intervals for tested contrasts.

Although an initial impression of the results from Experiment 1 may be that the performance difference between table-top and tablet interfaces is driven by a size difference, the complete

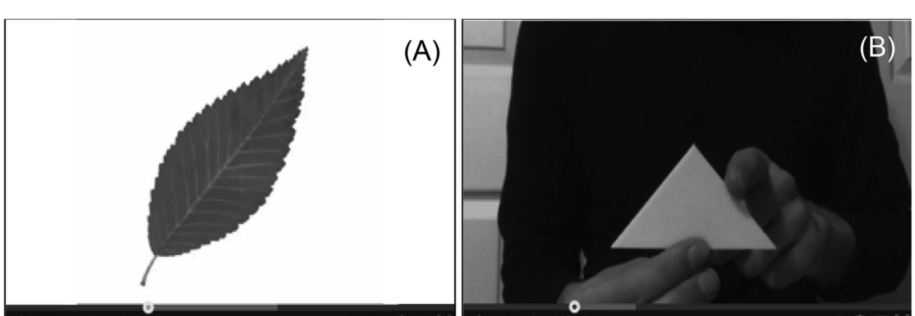

Fig. 3. (A) Screenshot of the Unprimed condition video that emphasized leaf characteristics. (B). Screenshot of the Priming condition video that emphasized tangram characteristics.

Table 3
Mean performance scores for tangram and math test tasks across Primed and Unprimed conditions.

	Unprimed		Primed		95% CI (Difference)
	M	*SD*	*M*	*SD*	
Puzzles solved	9.13	3.49	10.45	3.79	[−2.33, −0.3]*
Puzzle solution time (sec)	158.79	46.43	144.51	42.39	[1.55, 26.99]*
Math number correct	47.08	2.68	47.71	1.87	[−1.32, 0.06]
Math completion time (sec)	460.39	117.26	424.7	115.57	[2.77, 68.6]*

Note. CI = confidence interval.
*$p = 0.05$.

Table 4
Mean performance scores for tangram and math test tasks across Experiment 3 Unprimed tablet and Experiment 1 Table-top conditions.

	Unprimed tablet		Table-top		95% CI (Difference)
	M	*SD*	*M*	*SD*	
Puzzles solved	9.13	3.49	10.38	5.18	[−2.57, 0.08]
Puzzle solution time (sec)	158.79	46.43	140.08	70.72	[0.77, 36.65]*
Math number correct	47.08	2.68	47.42	2.17	[−1.1, 0.42]
Math completion time (sec)	460.39	117.26	394.83	141.97	[26.07, 105.05]*

Note. CI = confidence interval.
*$p = 0.05$.

Table 5
Mean performance scores for tangram and math test tasks across Experiment 3 Primed tablet and Experiment 1 Table-top conditions.

	Primed tablet		Table-top		95% CI (Difference)
	M	*SD*	*M*	*SD*	
Puzzles solved	10.45	3.79	10.38	5.18	[−1.16, 1.3]
Puzzle solution time (sec)	144.51	42.39	140.08	70.72	[−11.68, 20.56]
Math number correct	47.71	1.87	47.42	2.17	[−0.25, 0.83]
Math completion time (sec)	424.7	115.57	394.83	141.97	[−4.93, 64.68]

Note. CI = confidence interval.

alleviation of representational limitations through task-relevant priming in Experiment 3 clearly indicates that the performance difference in Experiment 1 is driven by representational rather than size differences. Additionally, this comparison provides an internal replication of the findings across samples, further reifying the conclusion that task representations are the driving force that causes performance differences across physical and digital interfaces.

6. General discussion

In three experiments, we showed that problem solving performance and subsequent effects on math performance are reliably influenced by the nature of the learning interface. Although previous research has presented similar findings, the present study is the first to assess representational differences in spatial task performance across physical and digital interfaces. Experiment 1 showed that the use of digital interfaces leads to poorer spatial task performance when compared with physical versions of the same task. A similar difference in performance was shown in a subsequent math test suggesting that the tablet condition led to less spatial priming than the tabletop condition. Despite an attempt at parity, the tablet interface involves a more constrained and less familiar interaction style compared to the tabletop task. However, in Experiment 2, participants who engaged in the tabletop paradigm with the constrained tablet interfaced performed comparably to those who engaged with the tabletop paradigm using the physical interactions typical for a physical task. This result shows that the performance difference reported in Experiment 1 could not be easily attributed to the physical limitations of the digital interface. Because tablets and other digital interfaces are becoming increasingly popular in education, despite lowered performance compared to physical versions, it is critical to understand this difference and develop practices to minimize it. Research from the literature on distributed cognition suggests that task performance is reliant on one's mental representation of the task space and includes the task goals, constraints, and the nature of various task interactions. In addition to explicit task instructions, this representation is also shaped by several task parameters including whether interaction with task-related symbols and objects use direct and familiar physical interaction or uses some more indirect (touch-screen, mouse, etc.) approach (Kirsh & Maglio, 1994; Zhang, 1997; Zhang & Norman, 1994). Predicting that the tablet task lacks sufficient cues to the 3D nature of the tangram pieces, as well as cues about the nature of physically manipulating them, Experiment 3 used video priming to boost participants' representation in the tablet paradigm. Results show that those with task-relevant pre-task priming performed both better on the puzzle task and better on the subsequent math test than controls who received task-irrelevant physical priming. Thus, Experiment 3 not only provides compelling evidence to show that differences in task representation drive performance decrements for digital interfaces, but also provides a simple and effective means to mitigate this difference.

The present results are consistent with several studies showing performance advantages for tangible interfaces over more indirect graphical user interfaces. However, little of this research has been geared specifically towards the impact of using of digital devices in educational contexts. Although the current study is the first to do so with spatial problem solving tasks, previous research has compared participants' reading comprehension when learning from tablets and e-readers rather than physical books. Although several studies show that physical texts are superior to digital ones (Ackerman & Goldsmith, 2011; Ackerman & Lauterman, 2012; Chen et al., 2014; Mangen & Kuiken, 2014; Mangen et al., 2013), others have failed to show consistent differences across various measures of comprehension (Eden & Eshet-Alkalai, 2013; Margolin, Driscoll, Toland, & Kegler, 2013, Subrahmanyam et al., 2013). Interestingly, the studies that fail to show this difference tend to use paradigms with relatively small amounts of text (e.g., Eden & Eshet-Alkalai, 2013; Margolin et al., 2013; Subrahmanyam et al., 2013). In contrast, comprehension differences were more frequently reported in studies that asked participants to read multiple lengthy passages (e.g., Ackerman & Goldsmith, 2011; Ackerman & Lauterman, 2012; Chen et al., 2014; Mangen & Kuiken, 2014; Mangen et al., 2013).

A limitation of the current study is in the age of the participants. Tangrams and similar tasks are commonly used to improve math concept formation and performance in younger children, yet participants in the current study were sampled from a college student population. College-aged students typically have several more years of experience solving spatial and math problems and thus may have produced different results than would be expected in younger samples. Some literature suggests that similar effects may be found for younger participants. For example, whereas the majority of tangible user interface research has been conducted with adults, the effectiveness of educational manipulatives has been found to vary across developmental stages (Carbonneau et al., 2013; Olkun, 2003; Wolfgang, Stannard, & Jones, 2003). However, it is possible that children who were introduced to touch-screen interfaces at an early age may not be susceptible to these performance differences, suggesting an effect of experience on one's ability to form adequate spatial representations. Discovery of this type of cohort effect would further underscore the importance of these current findings, particularly because students in underserved or underprivileged communities and schools may not have access to these touch-screen resources from an early age. Without this assumed level of experience, these children would be placed at a further disadvantage in situations where tablets are used in place of physical manipulatives. In light of this, we recommend that researchers explore how spatial and mathematical skills are affected by digital and physical interfaces for school-aged children in both short and long timeframes.

Similarly, research examining the effects of digital and physical books on children's reading comprehension is warranted because previous studies have also been almost exclusively conducted using adult participants. The majority of the existing research that addresses children's reading comprehension and listening compares physical text to enriched and interactive storybook software, thus confounding the comparison between physical and digital versions of text (see Zucker, Moody, & McKenna, 2009; for review). Findings from the limited number of studies that do compare digital and physical versions of the same text while controlling for enriched or interactive digital components indicate that young children do not demonstrate differences in reading comprehension (Jones & Brown, 2011) or listening (Roskos, Burstein, Shang, & Gray, 2014). However, more extensive research regarding children's reading comprehension for various interfaces may inform the future understanding of the effects of digitization on text processing and uncover potential cohort effects that are driven by early exposure to touch-screen interfaces.

Results from the present experiments do not necessarily suggest that the use of tablet and other digital devices be avoided in the teaching of spatial skills. Instead, by understanding the factors that influence performance differences across interfaces, we can suggest a set of techniques and practices for educators to use in order to minimize such differences in cases that lead to performance deficits. For example, Experiment 3 shows that educators can bootstrap performance in tasks involving digital interfaces with a video that primes and emphasizes the physical characteristics of the virtual manipulatives. This priming may enrich the learner's mental representation of the task and alleviate performance deficits between digital interfaces and their physical counterparts. Applying these techniques could allow educators to harness the benefits of new technology in the classroom while simultaneously minimizing any performance deficits that may result from the shift from physical to digital tasks. This is particularly important in cases where the knock-on educational benefit goes far beyond the spatial task itself, *and* where this benefit is dependent on one's performance in the primary task. Because tangrams and other manipulatives are used in service of improving or scaffolding the understanding of mathematical concepts, the preservation of the components of the task that ultimately lead to improved performance is of the highest importance.

Acknowledgments

Funding for this study was provided by a grant from the University of California Committee on Research.

References

Ackerman, R., & Goldsmith, M. (2011). Metacognitive regulation of text learning: on screen versus on paper. *Journal of Experimental Psychology: Applied, 17*(1), 18–32.

Ackerman, R., & Lauterman, T. (2012). Taking reading comprehension exams on screen or on paper? A metacognitive analysis of learning texts under time pressure. *Computers in Human Behavior, 28*(5), 1816–1828.

Brosterman, N. (1997). *Inventing kindergarten*. New York, NY: Henry N. Abrams.

Burnett, S. A., Lane, D. M., & Dratt, L. M. (1979). Spatial visualization and sex differences in quantitative ability. *Intelligence, 3*(4), 345–354.

Buzzetto-More, N., Guy, R., & Elobaid, M. (2007). Reading in a digital age: E-books are students ready for this learning object? *Interdisciplinary Journal of E-Learning and Learning Objects, 3*(1), 239–250.

Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology, 105*(2), 380–400.

Casey, M. B., Nuttall, R. L., & Pezaris, E. (2001). Spatial-mechanical reasoning skills versus mathematics self-confidence as mediators of gender differences on mathematics subtests using cross-national gender-based items. *Journal for Research in Mathematics Education*, 28–57.

Chen, G., Cheng, W., Chang, T. W., Zheng, X., & Huang, R. (2014). A comparison of reading comprehension across paper, computer screens, and tablets: does tablet familiarity matter? *Journal of Computers in Education, 1*(2–3), 213–225.

Cheng, Y. L., & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development, 15*(1), 2–11.

Delgado, A. R., & Prieto, G. (2004). Cognitive mediators and sex-related differences in mathematics. *Intelligence, 32*(1), 25–32.

Do-Lenh, S., Jermann, P., Cuendet, S., Zufferey, G., & Dillenbourg, P. (2010). Task performance vs. learning outcomes: a study of a tangible user interface in the classroom. In *5th European conference on technology enhanced learning*. Barcelona, Spain: Springer.

Eden, S., & Eshet-Alkalai, Y. (2013). The effect of format on performance: editing text in print versus digital formats. *British Journal of Educational Technology, 44*(5), 846–856.

Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods: an easy way to maximize the accuracy and power of your research. *American Psychologist, 63*(7), 591.

Esteves, A., van den Hoven, E., & Oakley, I. (2013). Physical games or digital games?: comparing support for mental projection in tangible and virtual representations of a problem-solving task. In *Proceedings of the 7th international conference on tangible, embedded and embodied interaction*. Barcelona, Spain: ACM.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191.

Fischer, U., Moeller, K., Bientzle, M., Cress, U., & Nuerk, H. C. (2011). Sensori-motor spatial training of number magnitude representation. *Psychonomic Bulletin & Review, 18*(1), 177–183.

Fitzmaurice, G. W., & Buxton, W. (1997). An empirical evaluation of graspable user interfaces: towards specialized, space-multiplexed input. In *Proceedings of the ACM SIGCHI conference on human factors in computing systems*. New York, New York: ACM.

Fjeld, M., & Barendregt, W. (2009). Epistemic action: a measure for cognitive

support in tangible user interfaces? *Behavior Research Methods, 41*(3), 876–881.
Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78*(4), 1343–1359.
Geary, D. C., Saults, S. J., Liu, F., & Hoard, M. K. (2000). Sex differences in spatial cognition, computational fluency, and arithmetical reasoning. *Journal of Experimental Child Psychology, 77*(4), 337–353.
Hawes, Z., Moss, J., Caswell, B., & Poliszczuk, D. (2015). Effects of mental rotation training on children's spatial and mathematics performance: a randomized controlled study. *Trends in Neuroscience and Education, 4*(3), 60–68.
Holmes, J., Adams, J. W., & Hamilton, C. J. (2008). The relationship between visuo-spatial sketchpad capacity and children's mathematical skills. *European Journal of Cognitive Psychology, 20*(2), 272–289.
Jacob, R. J., Ishii, H., Pangaro, G., & Patten, J. (2002). A tangible interface for organizing information using a grid. In *Proceedings of the SIGCHI conference on human factors in computing systems*. Minneapolis, Minnesota: ACM.
Jones, T., & Brown, C. (2011). Reading engagement: a comparison between e-books and traditional print books in an elementary classroom. *International Journal of Instruction, 4*(2), 5–22.
Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science, 18*(4), 513–549.
Kyttälä, M., Aunio, P., Lehto, J. E., Van Luit, J., & Hautamäki, J. (2003). Visuospatial working memory and early numeracy. *Educational and Child Psychology, 20*(3), 65–76.
Manches, A., O'Malley, C., & Benford, S. (2009). Physical manipulation: evaluating the potential for tangible designs. In *Proceedings of the 3rd international conference on tangible and embedded interaction*. Cambridge, UK: ACM.
Mangen, A., & Kuiken, D. (2014). Lost in an iPad: narrative engagement on paper and tablet. *Scientific Study of Literature, 4*(2), 150–177.
Mangen, A., Walgermo, B. R., & Brønnick, K. (2013). Reading linear texts on paper versus computer screen: effects on reading comprehension. *International Journal of Educational Research, 58*, 61–68.
Margolin, S. J., Driscoll, C., Toland, M. J., & Kegler, J. L. (2013). E-readers, computer screens, or paper: does reading comprehension change across media platforms? *Applied Cognitive Psychology, 27*(4), 512–519.
Marshall, P., Cheng, P. C. H., & Luckin, R. (2010). Tangibles in the balance: A discovery learning task with physical or graphical materials. In *Proceedings of the fourth international conference on tangible, embedded, and embodied interaction*. Cambridge, Massachusetts: ACM.
McKenzie, B., Bull, R., & Gray, C. (2003). The effects of phonological and visual-spatial interference on children's arithmetical performance. *Educational and Child Psychology, 20*(3), 93–108.
McLean, J. F., & Hitch, G. J. (1999). Working memory impairments in children with specific arithmetic learning difficulties. *Journal of Experimental Child Psychology, 74*(3), 240–260.
McNeil, N., & Jarvin, L. (2007). When theories don't add up: disentangling the manipulatives debate. *Theory into Practice, 46*(4), 309–316.
Moyer, P. S. (2001). Are we having fun yet? How teachers use manipulatives to teach mathematics. *Educational Studies in Mathematics, 47*(2), 175–197.
Olkun, S. (2003). Comparing computer versus concrete manipulatives in learning 2D geometry. *Journal of Computers in Mathematics and Science Teaching, 22*(1), 43–56.
Patten, J., & Ishii, H. (2000). A comparison of spatial organization strategies in graphical and tangible user interfaces. In *Proceedings of DARE 2000 on designing augmented reality environments*. New York, New York: ACM.
Project Tomorrow. (2014). The new digital learning playbook: understanding the spectrum of students' activities and aspirations. *Speak Up 2013 National Findings*. Retrieved from http://www.tomorrow.org/speakup/speakup_data_findings.html.
Rasmussen, C., & Bisanz, J. (2005). Representation and working memory in early arithmetic. *Journal of Experimental Child Psychology, 91*(2), 137–157.
Roskos, K., Burstein, K., Shang, Y., & Gray, E. (2014). Young children's engagement with e-books at school: does device matter? *SAGE Open, 4*(1), 1–9.
Schneider, B., Jermann, P., Zufferey, G., & Dillenbourg, P. (2011). Benefits of a tangible interface for collaborative learning and interaction. *IEEE Transactions on Learning Technologies, 4*(3), 222–232.
Spencer, C. (2006). Research on learners' preferences for reading from a printed text or from a computer screen. *International Journal of E-Learning & Distance Education, 21*(1), 33–50.
Stull, A. T., Barrett, T., & Hegarty, M. (2013). Usability of concrete and virtual models in chemistry instruction. *Computers in Human Behavior, 29*(6), 2546–2556.
Subrahmanyam, K., Michikyan, M., Clemmons, C., Carrillo, R., Uhls, Y. T., & Greenfield, P. M. (2013). Learning from paper, learning from screens: Impact of screen reading and multitasking conditions on reading and writing among college students. *International Journal of Cyber Behavior, Psychology and Learning, 3*(4), 1–27.
Terrenghi, L., Kirk, D., Sellen, A., & Izadi, S. (2007). Affordances for manipulation of physical versus digital media on interactive surfaces. In *Proceedings of the SIGCHI conference on human factors in computing systems*. San Jose, California: ACM.
Thompson, J. M., Nuerk, H. C., Moeller, K., & Kadosh, R. C. (2013). The link between mental rotation ability and basic numerical representations. *Acta Psychologica, 144*(2), 324–331.
Tuddenham, P., Kirk, D., & Izadi, S. (2010). Graspables revisited: multi-touch vs. tangible input for tabletop displays in acquisition and manipulation tasks. In *Proceedings of the SIGCHI conference on human factors in computing systems*. Atlanta, Georgia: ACM.
Wolfgang, C. H., Stannard, L. L., & Jones, I. (2001). Block play performance among preschoolers as a predictor of later school achievement in mathematics. *Journal of Research in Childhood Education, 15*(2), 173–180.
Wolfgang, C., Stannard, L., & Jones, I. (2003). Advanced constructional play with LEGOs among preschoolers as a predictor of later school achievement in mathematics. *Early Child Development and Care, 173*(5), 467–475.
Woody, W. D., Daniel, D. B., & Baker, C. A. (2010). E-books or textbooks: students prefer textbooks. *Computers & Education, 55*(3), 945–948.
Xie, L., Antle, A. N., & Motamedi, N. (2008). Are tangibles more fun? Comparing children's enjoyment and engagement using physical, graphical and tangible user interfaces. In *Proceedings of the 2nd international conference on tangible and embedded interaction*. Bonn, Germany: ACM.
Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science, 21*(2), 179–217.
Zhang, J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science, 18*(1), 87–122.
Zuckerman, O., & Gal-Oz, A. (2013). To TUI or not to TUI: evaluating performance and preference in tangible vs. graphical user interfaces. *International Journal of Human-Computer Studies, 71*(7), 803–820.
Zucker, T. A., Moody, A. K., & McKenna, M. C. (2009). The effects of electronic books on pre-kindergarten-to-grade 5 students' literacy and language outcomes: a research synthesis. *Journal of Educational Computing Research, 40*(1), 47–87.