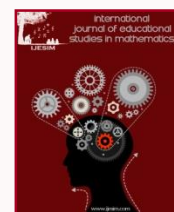




www.ijesim.com



Examining Patterns in Second Graders' Use of Virtual Manipulative Mathematics Apps through Heatmap Analysis

Patricia S. Moyer-Packenham¹, Stephen I. Tucker², Arla Westenskow³, Jürgen Symanzik⁴

^{1,2,3}School of Teacher Education and Leadership, College of Education and Human Services, Utah State University

⁴Department of Mathematics and Statistics, College of Science, Utah State University

ABSTRACT

This study examined 32 second-grade children's interactions with six virtual manipulative mathematics apps on the iPad using a novel analysis approach called heatmap analysis. Mathematical topics of the six apps focused on understanding skip counting and place value concepts. Children interacted individually with the apps under the guidance of an interviewer during 30-40 minute clinical interviews. The heatmap analysis revealed patterns in children's performance on the tasks and suggested individual cases for further analysis. For example, clusters of children showed high or low performance during the skip counting sequence. In the high-performing cluster, one child still struggled to monitor the skip counting process. In the low-performing cluster, one child showed growth on learning tasks but struggled on the assessments, which contained a different representation of the mathematical content and involved greater target numbers than in the learning app. There were fewer variations in place value task performance, but application of interaction techniques influenced children's growth on one learning app. Although some children showed varying performance when encountering challenging mathematics content in the higher levels of one app, other children inefficiently interacted with the app and did not reach the more challenging content

Keywords:

virtual manipulatives, heatmap, iPad apps, grade 2, touch screen devices

© 2015 IJESIM. All rights reserved

Article History:

Received 11.04.2015 Received in revised form 17.04.2015 Accepted 17.10.2015 Available online 14.12.2015

Introduction

"Apps" are technology applications for mobile devices with a touch screen (Gröger, Silcher, Westkämper, & Mitschang, 2013). Many mathematics apps contain a virtual manipulative (VM) that students can manipulate to support the visualization of mathematics concepts (Anderson-Pence, Moyer-Packenham, Westenskow, Shumway, & Jordan, 2014). Virtual manipulatives are defined by Moyer, Bolyard, and Spikell (2002) as "an interactive, Web-based visual representation of a dynamic object that presents opportunities for constructing mathematical knowledge" (p. 373). Although app use on touch-screen devices by young children is ubiquitous (Shuler, 2012), there is very little research on mathematics apps for children's learning. The purpose of this study was to extend this research base by examining patterns in children's learning when they used VM mathematics apps using an analysis approach called a heatmap analysis. This study was exploratory in nature and sought to explain patterns observed during children's interactions with the virtual manipulative mathematics iPad apps.

¹ Corresponding author's address: School of Teacher Education and Leadership, College of Education and Human Services, Utah State University

e-mail: patricia.moyer-packenham@usu.edu

DOI: <http://dx.doi.org/10.17278/ijesim.2015.02.004>

Theoretical Perspective

VM iPad apps have become popular because they use representations to display abstract mathematical concepts in a pictorial way and children can interact with these representations. Theorists like Bruner (1964) proposed that during cognitive development children make sense of their world through enactive means (i.e., the manipulation of physical objects) that then connect with iconic (visual images, pictures) and symbolic (words, numbers, symbols) representations. Representations include signs, symbols, models, images, or objects that stand for something else (Cai, 2005; Goldin & Shteingold, 2001) and are used to mediate and express learning. Representations help children make sense of mathematics and express their internal mental models of mathematics, and are therefore, important for children's cognitive development (Martin & Schwartz, 2005).

Studies have begun examining how physical and mental interaction happen simultaneously when children use touch-screen devices for mathematics learning (e.g., Moyer-Packenham, Anderson, et al., 2014; Moyer-Packenham et al., 2015; Moyer-Packenham, Westenskow, et al., 2014; Paek, 2012; Paek, Hoffman, & Black, 2013; Tucker & Moyer-Packenham, 2014; Tucker, Moyer-Packenham, Shumway, & Jordan, 2014). For example, Paek and colleagues (e.g., Paek, 2012; Paek et al., 2013; Paek, Hoffman, Saravanos, Kim, & Black, 2012; Paek, Hoffman, Saravanos, Black, & Kinzer, 2011) conducted a series of quantitative studies of a VM mathematics app and reported results on connecting feedback (i.e., audio and visual) and interaction types (i.e., touch versus mouse) in first- and second-grade students' learning. Generally, their studies suggested that audio feedback may be important initially, but interaction type (i.e., touch-screen vs. mouse) has more influence long-term. Other studies about the use of touch-screen devices are much more general in nature, and rather than examining features of the touch-screen environment, they focus simply on the use of the iPad for mathematics instruction in general. That is why there is a need for more exploratory studies that examine patterns and anomalies in children's interactions on touch-screen devices for mathematics learning.

Methods

This study was designed as an exploratory study to examine patterns in children's learning using several learning apps for mathematics using an approach called heatmap analysis. This type of approach can be useful to examining patterns that lead to the development of theory about the particular phenomenon under study. The research question that guided this study was: What are the patterns in children's learning when they complete a series of mathematical tasks using virtual manipulatives mathematics apps on a touchscreen device? The data that were gathered in this study were part of a larger funded project that included over 100 children ages 3-8. Publications from the larger project describe children's learning performance and efficiency, app affordances, and strategies children used when interacting with the apps (e.g., Moyer-Packenham, Anderson, et al., 2014; Moyer-Packenham, Bullock, et al., 2015; Moyer-Packenham, Shumway, et al., 2015).

Participants

The participants in this study were 32 children enrolled in Grade 2, ages 7 and 8. Children were recruited from elementary schools in Utah using informational brochures/letters. Most children were Caucasian (85%) and one-third of the children reported low Socio-Economic status (33%). Children's parents reported that there were touchscreen devices in most of their homes (91%) and that most children used a touchscreen device at least once per week (90%). Participants were assigned a code (from #69 to #100) for confidentiality and data analysis.

Procedures

To begin the study, researchers selected six VM iPad apps, designed interview protocols, and determined video data collection procedures. Researchers tested many apps with children in local schools prior to the study. The six apps were chosen because they were: appropriate for Grade 2, specific to the mathematics topics of skip counting and place value, and contained virtual manipulatives with affordances linked to positive learning outcomes (see e.g., those identified by Moyer-Packenham & Westenskow, 2013). The six apps were: Base-10 Blocks, Zoom Number Line, Place Value Cards, 100s Chart, Frog Number Line, and Counting Beads.

During the study, the children participated in one-to-one clinical interviews in a room equipped with a two-way mirror, an audio observer booth, and a built-in wall-mounted video camera. During the interviews,

researcher conducted the interview with the child while another researcher recorded observations from the observation booth. The interviewers all had elementary school teaching experience and expertise conducting mathematics clinical interviews.

At the beginning of each interview, the parent and child met the interviewer and observer and completed research paperwork. The parent went to the observation booth to watch the interview and the child went to the interview room with the interviewer. Each interview had two parts that focused on a different mathematics topic (see Table 1). Each child used six different apps during the interviews. During the first part of the interview, the child interacted with App#1 (100s Chart) as a pre-assessment of skip counting, then interacted with two learning apps (App #2 Frog Number Line and App #3 Counting Beads), and interacted with App#1 (100s Chart) again as a post-assessment. During the second part of the interview, the child interacted with App #4 (Base-10 Blocks) as a pre-assessment of place value, then interacted with two learning apps (App #5 Zoom Number Line and App #6 Place Value Cards), and interacted with App#4 (Base-10 Blocks) again as a post-assessment. The learning apps were presented in a random order (e.g., sometimes the Frog Number Line was presented first and sometimes the Counting Beads were presented first). The entire interview took 30-40 minutes.

Table 1: Sequence of the Interviews

Two-Part Interview	Interview Segment	Math Apps
Part 1 of the Interview Focused on Skip Counting	App #1 (pre-assess)	100s Chart
	App #2 (learning)	Frog Number Line
	App #3 (learning)	Counting Beads
	App #1 (post-assess)	100s Chart
Part 2 of the Interview Focused on Place Value	App #4 (pre-assess)	Base-10 Blocks
	App #5 (learning)	Zoom Number Line
	App #6 (learning)	Place Value Cards
	App #4 (post-assess)	Base-10 Blocks

Note. Information about each app is located in the appendix.

Data Sources

The primary instruments used to collect data during the study were video recordings (using a wall-mounted camera and a wearable GoPro camera) and observer notes, providing multiple observational perspectives (Roschelle, 2000). The wall-mounted video camera was positioned six feet away from children and captured all actions and interactions with the interviewer. The wearable GoPro camera captured an up-close perspective of the touchscreen device and the child's hands interacting with the virtual manipulatives. It was positioned six inches away from the iPad (see Figure 1). Observers were six feet away from children behind two-way glass. All session videos were digitized and accessed through a password-protected site.

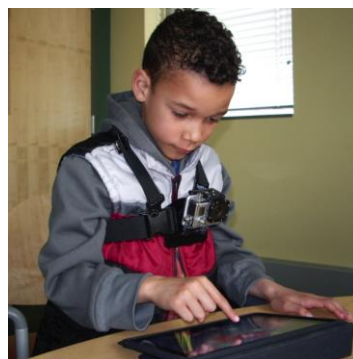


Figure 1. Child with wearable GoPro camera.

Data Analysis

Smith et al. (2006) define learning progressions as "descriptions of successively more sophisticated ways of reasoning within a content domain based on research syntheses and conceptual analyses" (p. 2). In this study we use the term "learning progressions" because we focus on evidence of reasoning about a specific

content domain (i.e., skip counting and place value) during a specific activity (i.e., completing tasks while interacting with a particular app). To analyze the video data we created a different scoring rubric for each set of tasks. Each scoring rubric was based on a hypothesized learning progression of levels of reasoning about the content (e.g., skip counting or place value) that students might exhibit in the context of interacting with the app to complete the assigned tasks. For example, there were seven possible levels of reasoning about skip counting that children could exhibit while completing tasks by interacting with the 100s Chart app (see Table 2). Higher scores indicated higher levels of reasoning. A child who guessed without counting received a level score of 1 for that task, while a child who skip counted correctly the entire sequence received a level score of 7 for that task.

Table 2: Learning Progression Scoring Rubric for Grade 2-100s Chart (Pre and Post Assessment)

Level	Mathematical Learning Progression Expectations; Growth in Skip Counting Abilities
1	Child guesses without counting; no response
2	Correct iterative counting with no pattern to skip counting intervals
3	Counting on from first number with correct intervals (using fingers or objects to count intervals - dual counting)
4	Skip counts first two or more sequence parts and counts on for subsequent parts
5	Skip counts with minor error (one less or one too many in one sequence)
6	Skip counts correctly by adding (verbal or other signs indicating calculation)
7	Skip counts correctly entire sequence

Researchers coded the video data for each child and each task attempt by assigning a Learning Progression Score Level to each attempt during the interviews. After coding all of the video data, the Learning Progression Score Levels were then converted to a scale score with $1/\text{maximum possible score}$ reflecting the least sophistication and $\text{maximum possible score}/\text{maximum possible score} (= 1)$ reflecting the most sophistication on each of the learning progressions. For example, in Table 2, a child who scored level 3 on the first task, level 5 on the second task, and level 5 on the third task for the 100s Chart app would have scale scores of $3 \div 7 = 0.43$, $5 \div 7 = 0.71$, and $5 \div 7 = 0.71$. Scaling the data allowed for comparisons across the sequences. To ensure internal validity, 20% of the videos were double coded by two researchers.

Heatmap Analysis

Heatmaps are a popular graphical way to summarize data, observe relationships among several statistical variables (the columns in a heatmap), and organize the observations from numerous participants (the rows in a heatmap) – all in one single graph. Heatmaps have been made widely popular by Eisen et al. (1998). But, as Wilkinson & Friendly (2009) pointed out, heatmaps and related graphs can be dated back to the 19th century. The underlying idea for heatmaps is that the data are split into different intervals that are assigned to a color. For our heatmaps, we broke the standardized data into ten intervals of width 0.1 and used a blue/red divergent color scheme (see Figure 3). The darker the blue, the closer the values were to 0. The darker the red, the closer the values were to 1.0. Light blue and light red colors represent values around 0.5. When children completed different numbers of tasks on an app, we used black to represent missing observations (i.e., where some children did not attempt as many tasks as other children), assigning them the value -0.1. The exact colors and underlying numerical intervals can be found in the upper left corner of the heatmap plots. A histogram is overlaid on this color representation to show how many times values from each interval occur overall (i.e., in all rows or columns combined) in a heatmap.

The rows and columns in a heatmap can be presented sequentially or may be rearranged so that rows (or columns) that are similar to each other are placed next to each other in the heatmap representation. In the heatmaps we included in this report, we always sort the rows, which represent individual children. At times, we also sort the columns, which represent the variables in the study. When we expected that the data might reveal some learning/improvement over time, we left the columns in their original (temporal) order.

The similarity of rows (or columns) is determined via *hierarchical clustering* (see Section 14.3.12 in Hastie, Tibshirani & Friedman, 2001, for a comprehensive overview). The results of this clustering are shown via *dendrograms* to the left of the heatmap (for the rows) and, if used, on top of the heatmap (for the columns). A dendrogram is an easily interpretable tree-based representation that displays the arrangements of the clusters from a hierarchical cluster analysis. The terminal nodes in a dendrogram represent individual observations. The inner non-terminal nodes represent groups of at least two observations. The height of each inner node is

an indicator of how similar (or dissimilar) the observations are in the branches that lead towards the terminal nodes. If two (or more) individual observations are connected at height zero (i.e., directly at the height of the terminal nodes), this implies that all numerical values are exactly the same for these observations. (See also Section 14.3.12 in Hastie, et al., 2001). These techniques have been used in a few other educational studies before. Bowers (2010) used heatmaps to display the results from a hierarchical cluster analysis of a longitudinal K-12 grading history data set and Kinnebrew, Segedy, and Biswas (2014) used heatmaps to display the temporal development of students' behaviors in learning environments.

Results

The results are presented in three parts: Part 1 – learning progression scores from the Skip Counting sequence; Part 2-learning progression scores from the Place Value sequence; and Part 3-learning progression scores from the whole interview. Each section describes the app, the tasks, and the results for all children and for individual cases. This highlights patterns in children's learning progressions. The results are reported for children who interacted with all apps in a given sequence (e.g., there are 27 children in Part 1, 26 children in Part 2, and 25 children in Part 3).

Part 1: Skip Counting Sequence

The Skip Counting sequence included tasks using the 100s Chart, Counting Beads, and Frog Number Line apps.

1.a. 100s Chart. The 100s Chart app was a digital version of a hundreds chart (see Figure 2). The pre- and post-assessments on this app asked children to count by 4 to 28, 6 to 18, and 9 to 36, and to tell how many times they counted by the given number to reach the target (e.g., "I counted by six three times to get to 18"). The app displayed the numbers in ten rows of 10, from 1-100. Children could track their skip counting because tapping spaces on the board made the number turn red; a second tap caused the number to disappear, and a third tap caused the number to reappear in black.

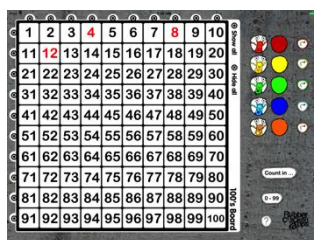


Figure 2. Screenshot from 100s Chart app.

Figure 3 shows a heatmap of the combined pre- and post-assessment results for the 26 children that completed the Skip Counting sequence. Overall, no children had exactly the same outcomes on the 100s Chart app. However, in the dendrogram on top of the heatmap, three pairs of two similar variables can be seen: Pre.Skip.by.4 and Post.Skip.by.4, Pre.Skip.by.6 and Pre.Skip.by.9, and Post.Skip.by.6 and Post.Skip.by.9. This means that the pre- and post- results for the skip by four task were similar, pre-assessment results for skip by six and by nine were similar, and the post-assessment results for skip by six and by nine were similar.

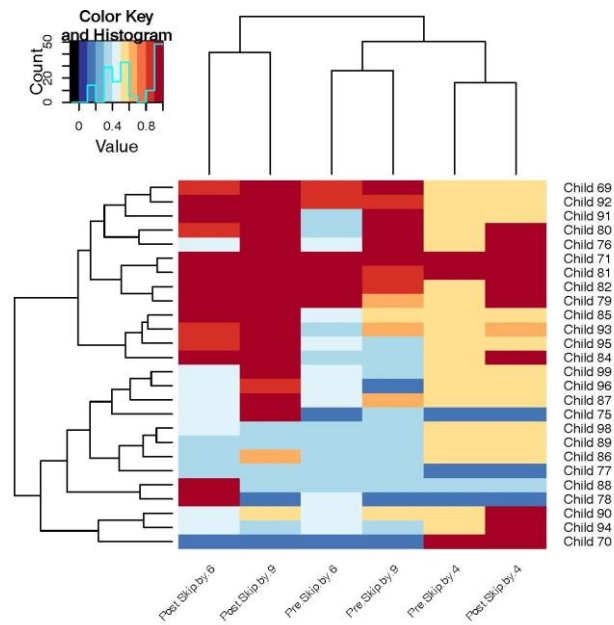


Figure 3. Heatmap of Skip Counting Pre/Post Combined 100s Chart.

There was a relatively even distribution of areas of growth on the Skip Counting Pre/Post tasks. Some showed growth on all three tasks, only on 4’s tasks, only on 6’s tasks, only on 4’s and 6’s tasks, only on 6’s and 9’s tasks, only on 4’s and 9’s tasks, and only on 9’s tasks. The learning apps appeared to promote learning of skip counting, although learning was in a variety of tasks across the children. When children had difficulties skip counting by 4’s, the analysis indicated that many children reached 12 or 16 before making a mistake. Additional patterns that align with other apps are discussed later in the combined sections (see 1.c and 1.e).

1.b. Counting Beads. The Counting Beads learning app (see Figure 4) presented a virtual tray containing groups of linked beads in strings of one to ten and a mat. Children were asked to make bead strings to count by 4 to 28, 6 to 18, and 9 to 36, and to tell how many times they counted by the given number to reach the target. The app allowed the child to place one type of string per task (e.g., only strings of five beads). Every time a child placed a string of beads, a number card (indicating the cumulative number of beads on the mat) appeared in a pile in the last tray section. The child could move the number cards to label the beads at any time.



Figure 4. Screenshot from Counting Beads app.

Figure 5 shows a heatmap of the Counting Beads learning app results for the 26 children that completed the Skip Counting sequence. In the dendrogram on top of the heatmap, there are three groups of two variables: Skip.by.6 and Skip.by.9, Skip.by.4 and Question.4, and Question.6 and Question.9, indicating similar outcomes within each pairing. There are three main clusters of children. In Cluster 1 (10 children: #82 to #90), children had medium to high scores for most of the variables. In Cluster 2 (7 children: #94 to #80), children had medium scores for most of the variables. In Cluster 3 (9 children: #75 to #86), children had lower scores for several of the variables. This indicates a fairly even distribution of relatively high, medium, and low performance. Because this was the “learning app” portion of the interview, it makes sense that children would exhibit a variety of learning progression score as they were developing understanding on the mathematics topic.

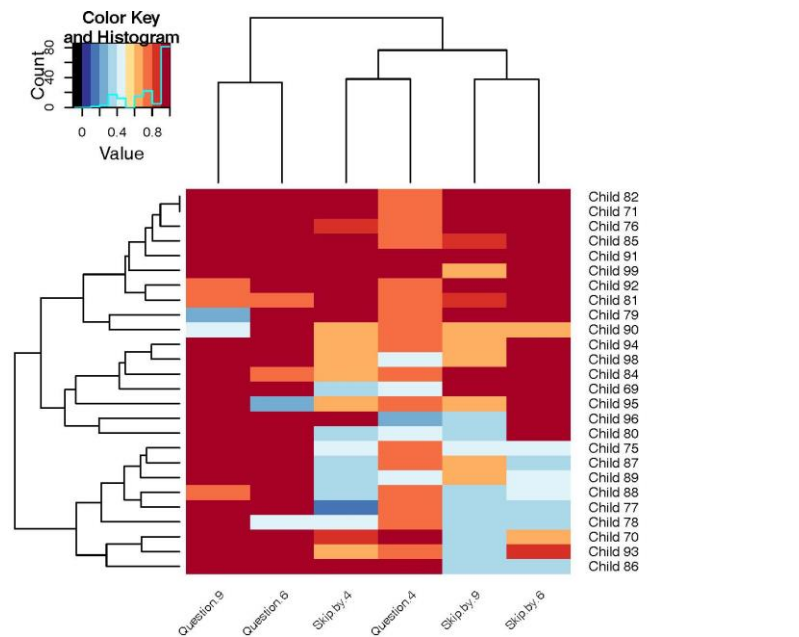


Figure 5. Heatmap of Skip Counting Learning App Counting Beads.

The main difference between clusters of children occurred during the modeling tasks (as opposed to the follow-up question). Children in Cluster 1 generally had high scores on all three tasks. Children in Cluster 2 generally had high scores when counting by 6, medium or high scores when counting by 9, and medium scores when counting by 4. Children in Cluster 3 generally had low scores on all three tasks. There was less variance among groups on the follow-up questions. This may be because a child who misrepresented the skip counting sequence (e.g., placed five sets of nine instead of four sets) could still correctly answer the follow-up question by counting the number of bead groups used to reach the target number.

However, the task involving modeling skip counting by four to 28 required counting more times (seven times) than skip counting by six to 18 (three times) or skip counting by nine to 36 (four times). Counting the number of bead groups may have been easier when there were fewer groups to label and count. Furthermore, the app scaled bead sizes by the number of beads in the string (e.g., in a set of two beads, each bead was larger than each bead in a set of five beads). Many children began their sequences near the middle of the screen. This sometimes meant the representation of seven sets of four beads to make 28 beads extended beyond the edge of the display screen, whereas the other two tasks usually fit within the screen.

1.c. Heatmap analysis for Pre/Post 100s Chart and Counting Beads apps combined. Figure 6 shows a heatmap of the combined 100s Chart pre-/post-assessment app, and the Counting Beads app results for the 26 children that completed the Skip Counting sequence. In the dendrogram on top of the heatmap, there are three groups of two variables: Skip.by.6 and Skip.by.9, Skip.by.4 and Question.4, and Question.6 and Question.9, indicating similar outcomes within each pairing. There were three main clusters of children: In Cluster 1 (9 children: #92 to #79), children had medium to high scores for most of the variables. In Cluster 2 (6 children: #84 to #93), children had a few medium scores for some of the variables. In Cluster 3 (11 children: #94 to #78), children had lower or medium scores for several of the variables.

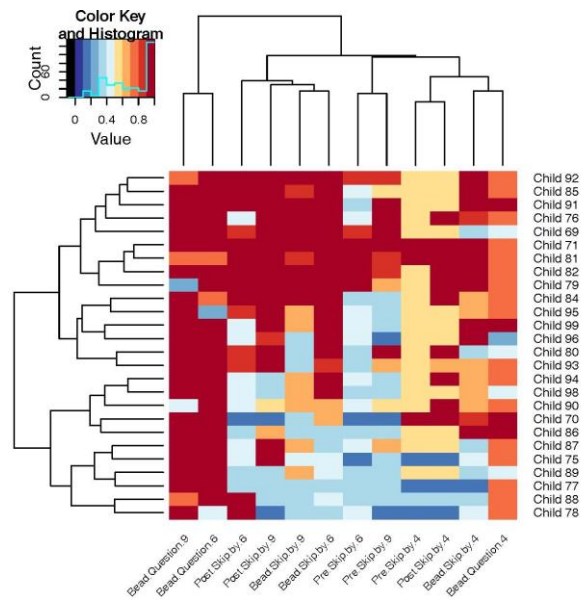


Figure 6. Heatmap of Pre/Post 100s Chart and Counting Beads Apps Combined.

The clustering of the tasks shows relationships among children’s performance across these apps. The first group of tasks may relate to the previously mentioned design of the beads app features. The second group of tasks provided additional exposure to a similar (or identical) task as the pre-assessment. The third group of tasks were the first exposure to the tasks and had somewhat lower overall performance. The fourth group of tasks all involved skip counting by 4 and scores were generally high, leaving less room for growth. Children in Clusters 1 and 2 generally scored highly on the post-assessment skip by 6 and 9 tasks, but Cluster 1 scored higher on the pre-assessment tasks. On the pre-assessment, four children in these clusters correctly used patterns on the hundred chart (indicative of the highest learning progression level) during the skip count by 9 task; this increased to 12 children on the post-assessment task. Children in Cluster 3 generally scored lower than Clusters 1 and 2 on the post-assessment skip by 6 and 9 tasks. Children who achieved a perfect score on the post-assessment skip by 9 task correctly used patterns on the hundred chart.

1.d. Heatmap analysis for Frog Number Line. Children responded to app-generated prompts in the Frog Number Line learning app (see Figure 7) by hopping a frog across a number line labeled from 0-15 by intervals of one. The app displayed randomly sequenced instructions to skip count by 2’s, 3’s or 4’s, and a frog appeared on the number line at the starting position of 0, 1, 2 or 3. To indicate a response, the child let go of the frog over a number, highlighted in yellow to track placement. The label turned green if the child was correct and displayed a positive message (e.g., “Winning!”); then a new task appeared.

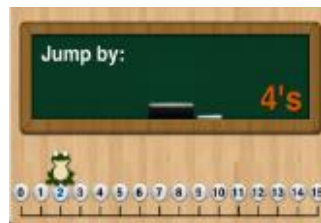


Figure 7. Screenshot of Frog Number Line app.

Figure 8 shows a heatmap of the Frog Number Line results for the 26 children that completed the Skip Counting sequence. Children completed a varied number of tasks (4-16) and all tasks had the same highest possible learning progression score. Four children (#81, #69, #82, and #71) achieved perfect scores and two children (#90 and #80) only scored below 1.0 once each. Eight of the 26 children had scores below 0.5 in their first four tasks (T1-T4), and then improved on subsequent tasks. Some children had little room for growth because they achieved high levels initially.

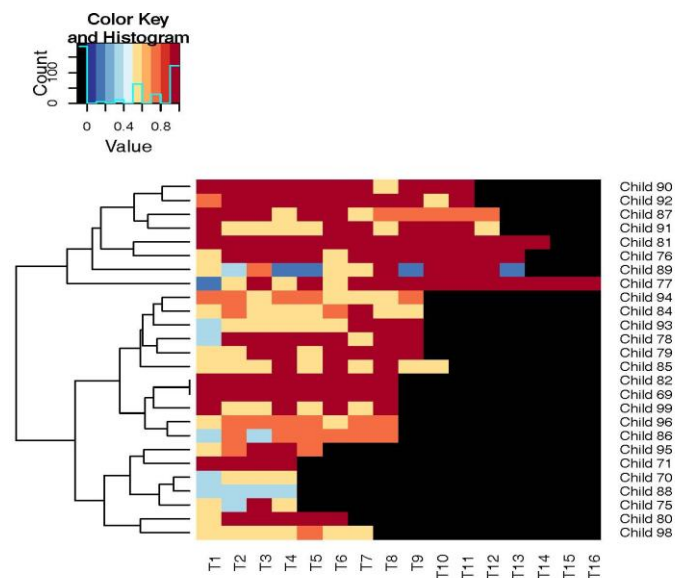


Figure 8. Heatmap of Skip Counting Learning App, Frog Number Line.

Two examples of children whose score increased during their use of the app were #77 and #93. Child #77 began on T1 (task 1) at a lower level on the learning progression (blue), alternated between medium and high levels for T2-T6 (yellow and red), and then consistently attained the highest level of the learning progression on T7-T16 (red). Child #77 initially had difficulty on some tasks that required starting at a number other than zero. After a period of counting on from the starting number, Child #77 began skip counting correctly by two, three, or four from any starting number. Child #93 was first prompted to count by threes starting from two, but first counted to three and attempted to continue counting by three. After being shown to count by the prompt number from the place where the frog started, Child #93 correctly completed this task and the following five tasks using iterative counting, beginning at the starting number and counting on to reach the subsequent numbers. On the final three tasks, Child #93 was able to correctly skip count the entire sequence. This demonstrates gradual change and shifts in children's learning from lower levels of mathematical reasoning to higher levels of mathematical reasoning during these tasks.

Other children did not increase their scores on these tasks. For example, Child #88 achieved low scores throughout the tasks, pointing to each number and counting by ones, moving the frog only after identifying the next number in the sequence. Some children performed differently depending on the mathematical content of the task. Child #99 scored higher on tasks requiring skip counting by two than on tasks requiring skip counting by three or four. Child #75 scored higher when beginning from 0 than other numbers. This indicates that closer examination of children's mathematical interactions with virtual manipulative mathematics iPad apps may reveal nuances in the trajectory of their mathematical understandings (Clements & Sarama, 2009); Sarama & Clements, 2009).

1.e. Analysis of patterns across the skip counting app sequence. Child #71 earned the highest possible score on every skip counting task except the follow-up question for the skip by four tasks in Counting Beads, during which Child #71 counted the groups to determine how many it took to reach the target number (i.e. the second-highest possible score). Similarly, Child #81 could already skip count well, but was relatively less adept at monitoring how many times one counts by a given number to reach a target number. These children demonstrate the incremental steps in the development of a more complete understanding of a mathematical topic.

Child #77 earned very low scores on all pre-assessment, post-assessment, and Counting Beads tasks requiring modeling skip counting, but was able to identify the number of times counted to reach the target number in Counting Beads. Child #77 made growth on Frog Number Line tasks, which required skip counting on a number line by two, three, or four. Therefore, Child #77 improved in skip counting by two, three, and four (to numbers no greater than 15) using a number line representation, but had not transferred this skip counting skill to the hundreds board representation beyond 15. Although Child #77 could determine the number of times counted to reach a target number, representing skip counting to target numbers beyond 15

remained difficult. Overall, these results suggest that in-depth analysis of children’s interactions with virtual manipulative mathematics iPad apps may provide insights into children’s developing mathematical understandings.

Part 2: Place Value Sequence

The Place Value sequence included tasks using the Base-10 Blocks, Zoom Number Line, and Place Value Cards apps.

2.a. Base-10 Blocks. For the pre- and post-assessment on the Base-10 Blocks app (see Figure 9), children used the blocks to complete six tasks modeling specific quantities or quantities within a given range. The app presented children with a place value chart and blocks (ones, tens, and hundreds) to use to form the target quantity. When the child altered the number of blocks, the number tiles changed accordingly. The app stacked the blocks and bundled groups of ten blocks, moving the new block into the appropriate space on the place value chart.



Figure 9. Screenshot of Base-10 Blocks app.

Figure 10 shows a heatmap of the combined pre- and post-assessment results for the 27 children that completed the Place Value sequence. Most children had perfect scores on the pre- and post-assessment using the Base-10 Blocks app. Seventeen of the 27 children had identical outcomes with a perfect score of 1.0 for all twelve tasks. Only on variables Pre4, Post4, Pre6, and Post6 did more than one child score less than 1.0. These generally high scores resulted in little variation. The video data revealed that the only instances where children made errors on the tasks for this app were when they did not follow instructions for the task (e.g., choosing to model a different number) or made mathematical errors (e.g., counting incorrectly to find 20 less than 267). The assigned tasks were too easy for this group of students. This is in contrast to pilot testing of this app where there was quite a bit of variation in children’s performance.

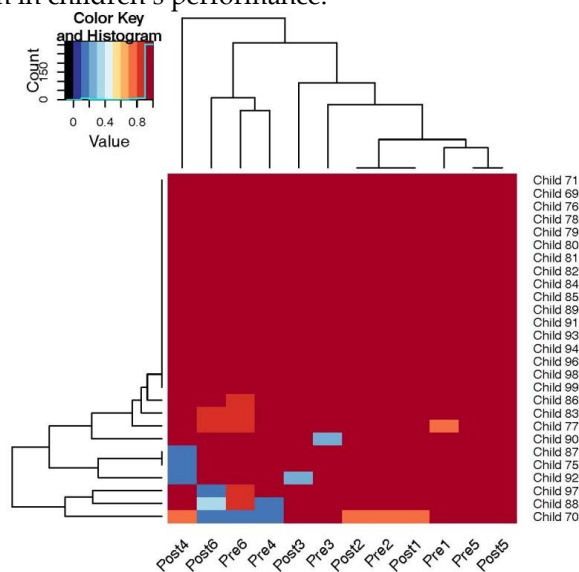


Figure 10. Heatmap of Place Value Pre/Post Combined from Base-10 Blocks.

2.b. Heatmap analysis for Zoom Number Line. In the Zoom Number Line learning app (see Figure 11), children used multi-touch input to navigate a number line and place app-generated target numbers. To place

the number, children swiped to move the number line left or right, or zoomed in or out to change interval size (e.g., ones, tens, hundreds) using a pinching gesture, before tapping to pop the bubble and place the number.



Figure 11. Screenshot from Zoom Number Line app.

Figure 12 shows a heatmap of the Zoom Number Line results for the 27 children that completed the Place Value sequence. There was no standard level of task difficulty, so scores for each task were scaled based on the highest possible learning progression rubric score to facilitate comparison. Children completed between 25 and 46 tasks using the Zoom Number Line app. Black areas on the heatmap show that children completed different numbers of tasks for this app. There were approximately 14 tasks per level, and many children appeared to improve around tasks 14, 28, or 42. The most common score, 0.4, indicated that students demonstrated knowledge of one- and two-digit numbers on a number line. Scores generally increased during level four, wherein tasks offered children the opportunity to score higher on the Learning Progression Scoring Rubric (e.g., involving three-digit numbers or changing intervals). The zoom function was necessary for some tasks in and beyond level four but not required before level four. Most of the variations in the heatmap were due to the mathematical requirements of the tasks the children encountered, while the number of tasks also influenced clustering.

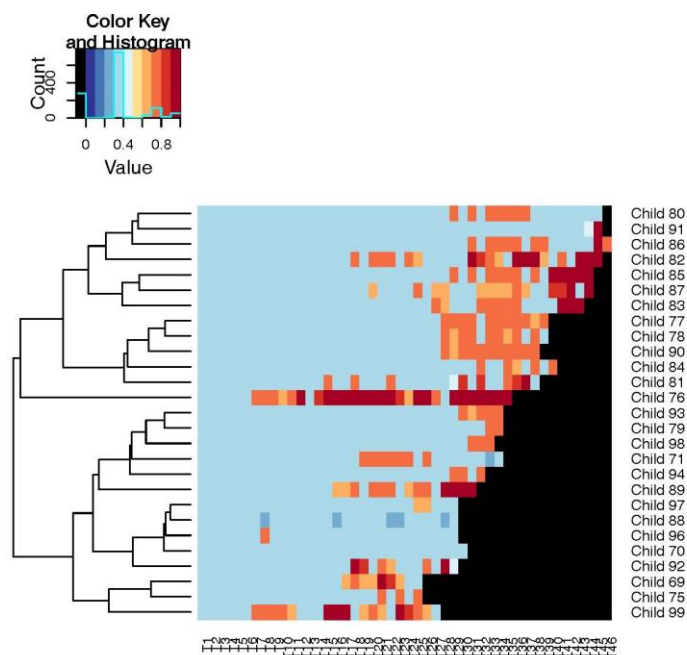


Figure 12. Heatmap of Place Value Learning App Zoom Number Line.

The Zoom Number Line task difficulty varied by app level, and some tasks were not at a high learning progression level. For example, the heatmap shows that Child #82's scores increased over time. Examining the video revealed that Child #82 often earned the highest possible scores until reaching different mathematics content on level 5, wherein the child initially struggled. By the end of level 5 and throughout level 6, Child #82 once again scored the highest possible scores. This indicated that Child #82 made growth in mathematical understanding within the levels of the app. Child #83 also demonstrated growth in mathematical

understanding during level 4 once the child chose to change intervals via zooming. However, Child #99 struggled with the mathematics content on level 5 and did not achieve the highest possible scores, while Child #76 achieved the highest possible score for 30 of 35 tasks. Child #88 avoided changing intervals and at times explored the swipe function by rapidly increasing or decreasing as far as allowed by the app before trying to find the target number. Children #70, #96, and #91 encountered few or no tasks that required changing intervals and avoided zooming when it was not required.

1.d. Place Value Cards. In the Place Value Cards learning app (see Figure 13), children represented three-digit target numbers without zeros (e.g., 298, not 201) using single touch input to drag place value cards. The target numbers appeared in word form and were read aloud by the app. Columns of cards were arranged left to right by descending place values, and within each column, cards were arranged from least to greatest. Children could use the addition frame to represent the number (e.g., $900 + 40 + 3$) and then move the cards to the “sum” box, or move the cards directly to the sum box.



Figure 13. Screenshot of Place Value Cards app.

Figure 14 shows a heatmap of the Place Value Cards results for the 27 children that completed the Place Value sequence. Children completed between 4-29 tasks. Most children scored 1.0 on every attempted task; only two children (#70 and #80) had more than one score below 1.0. Although the number of tasks influenced clustering, virtually no pattern is evident. The video data indicated that the only mistakes were made when trying to represent irregular numbers (i.e., numbers ending in eleven, twelve, or teens). As with the Base-10 Blocks results, there was little variation, with children achieving high accuracy on most tasks.

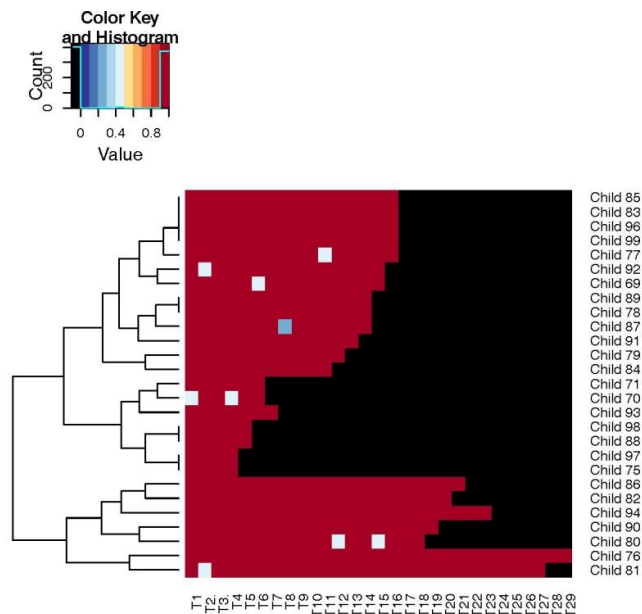


Figure 14. Heatmap of Place Value Learning App, Place Value Cards.

Part 3: Analysis of the Entire Interview

Figure 15 shows a heatmap of the combined pre- and post-assessment results for the 25 children that completed both the Place Value and Skip Counting sequences of the interview.

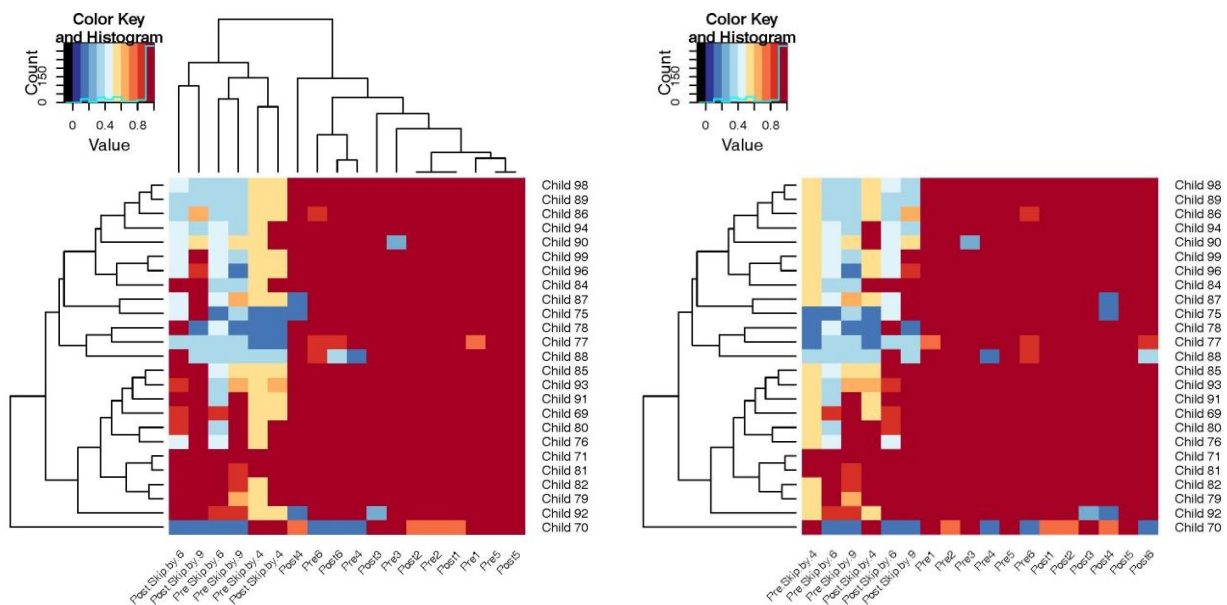


Figure 15. Heatmap of Pre/Post for the Entire Interview – Place Value and Skip Counting. Columns are sorted based on similarity (left) and kept in sequential order of the interview (right). The ordering of the rows does not change.

Overall, there were no children with exactly the same outcomes. In the dendrogram on top of the left heatmap, there are two main groups of variables: Pre/Post 1 through 6 (which is the group of variables for the Place Value sequence) and Pre/Post skip.by 4, 6, and 9 (which is the group of variables for the Skip Counting sequence). As the heatmap shows, the outcomes for the Place Value and Skip Counting Pre/Post Sequences are quite different.

There are two main clusters of children: In Cluster 1 (13 children: #98 to #88 on the right), children had low to medium scores for some of the variables. In Cluster 2 (11 children: #85 to #92 on the right), children had medium to high scores for most of the variables. Child #70 differs considerably from both main clusters of children, with relatively lower scores overall. The clustering of children for the overall interview results almost perfectly matches the higher and lower clusters from the Pre/Post combined clustering of the 100s Chart app. Child #84 is the exception, with high scores on Place Value Pre/Post tasks and had relatively low scores for Pre.Skip.by.6 and Pre.Skip.by.9 of Skip Counting Pre/Post, thus placing the child into the “weaker” group overall. Together, the grouping and clustering suggest that results from each set of tasks was more similar to results on other tasks in the same sequence than to results from tasks in the other sequence.

Discussion

In this study we used data visualization methods (heatmaps) that are less common in the field of education to examine educational data about children’s experiences with virtual manipulatives mathematics iPad apps. The visualization of the clustered data via dendrograms and heatmaps allowed us to explore children’s learning during their mathematical experiences with six iPad apps. This type of visualization was useful to our research in several ways. First, it created visual examples of children’s growth over the course of the interview; second, it identified when interview protocols were too easy for children; and, third, it identified areas to look to for further research and the generation of questions about the use of virtual manipulatives mathematics apps on iPads.

When we compare the two sequences of the interview (i.e., one sequence focused on skip counting and one sequence focused on place value), we see very different results in the heatmaps for this group of second-grade children. This is particularly evident in Figure 15 where we compare the two parts of the interview in the same heatmap. We see a great deal of variation in the Skip Counting sequence of the interview, while there is very little variation in the Place Value sequence of the interview. A close look at the Skip Counting sequence shows that the pre-assessment app, one of the learning apps (Counting Beads), and the post-assessment app all focused on skip counting by 4s, 6s, and 9s using different representations. Additionally, the other learning app in this sequence (Frog Number Line) allowed children to practice their skill of skip counting. However, in the Place Value sequence of the interview, the tasks in the two learning apps did not closely resemble the pre-assessment and post-assessment. Furthermore, the Place Value pre- and post-assessment seemed to be too easy for this group of second-grade children. While these Place Value apps were tested with second-grade

children in a pilot and the task difficulty was high for the pilot children, it turned out to be too easy for the children who participated in our study.

By examining the videos of individual children we observed that they were more likely to use patterns to skip count by 9 on the post-assessment. This strategy was effective for increasing their accuracy on the task. Those who completed this task correctly on the pre-assessment at a low level on the learning progression may have learned from the pattern they saw on the 100s Chart and then completed the task at a higher level on the learning progression during the post-assessment tasks. Children who performed highly on the post-assessment skip by 9 (many from Clusters 1 & 2; both correct in Cluster 3) often used patterns on the hundred chart for every part of this task (i.e., 9, 18, 27, 36). Lower-performing children may have used these patterns less effectively (i.e., count to 9, count to 18, pattern for 27 & 36).

The heatmaps analysis helps us to see relationships among the different tasks that were presented to the children in the study. For example, we can see that there is a relationship between the Counting Beads tasks and the 100s Chart pre-post tasks for skipping by 6 and 9. Learning what questions are related to each other for the children helps us to think about how we might present and teach mathematics topics and content to capitalize on natural relationships among those topics that might help children to better learn and understand mathematics topics.

When we examine individual instances of children by returning to the videos, we also see that there were instances of growth within and across the apps for some children. In particular, there were children who showed growth across the Skip Counting sequence of the interview. However, there were also children, like Child #82 and #83 who showed growth within the Zoom Number Line app. This growth within a single app was also true of the Frog Number Line app where we saw gradual change from lower levels to higher levels of sophistication during children's interactions with this app. This shows how the data visualization using the heatmap analysis process can help us to view the whole group of children as well as to view individual children and their growth.

Overall, these results suggest that analysis of children's interactions with virtual manipulative mathematics iPad apps using heatmaps and detailed review of the interactions may provide multiple levels of insight about children's mathematical understandings. It is important to recognize that many factors may influence the clusterings and groupings in the heatmaps. For example, the mathematical model influences the display of the data (e.g., assigning a value of -0.1 to "missing" data produces different clustering than assigning the average score for the "missing" task). Furthermore, the nature of the data, such as using learning progressions based on reasoning as opposed to dichotomous "correct/incorrect" scoring influences the conclusions one can draw from the data. Thus, one should seek underlying explanations for the patterns. Combining these levels of analysis may allow for more thorough interpretation of children's mathematical understandings than using either technique alone.

These data were part of a larger study that also included preschool and kindergarten children also using virtual manipulatives mathematics learning apps on iPads. Our future research will include examining the data from the larger study to conduct similar heatmap analyses.

Conclusion

Using this method of data visualization of the clustered data via dendrograms and heatmaps helps us to display and better understand complex educational data about children's learning gains during mathematical experiences. This type of visualization allows a teacher to ask/answer questions like - Are there skills that are related to each other? Which children perform similarly well? What can we tell about our children from looking at these visualizations? Which children show improvement in which category for which task or at which point in time? Wilkinson & Friendly (2009) stated: "The cluster heat map is well known in the natural sciences and is one of the most widely used graphs in the biological sciences" (p. 180). At some point in the future, it may play an equally important role for the display of educational data.

References

- Anderson-Pence, K. L., Moyer-Packenham, P. S., Westenskow, A., Shumway, J., & Jordan, K. (2014). Relationships between visual static models and students' written solutions to fraction tasks. *International Journal for Mathematics Teaching and Learning*, 15, 1-18.
- Bowers, A. J. (2010). Analyzing the longitudinal K-12 grading histories of entire cohorts of students: Grades, data driven decision making, dropping out and hierarchical cluster analysis. *Practical Assessment, Research & Evaluation*, 15(7). Available online: <http://pareonline.net/getvn.asp?v=15&n=7>.
- Bruner, J. S. (1964). The course of cognitive growth. *American Psychologist*, 19(1), 1-15.
- Cai, J. (2005). U.S. and Chinese teachers' constructing, knowing, and evaluating representations to teach mathematics. *Mathematical Thinking and Learning*, 7(2), 135-169.
- Clements, D. H., & Sarama, J. (2009). *Learning and teaching early math: The learning trajectories approach*. New York: Routledge.
- Eisen, M.B., Spellman, P.T., Brown, P.O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 95(25), 14863-14868.
- Gröger, C., Silcher, S., Westkämper, E., & Mitschang, B. (2013). Leveraging apps in manufacturing. A framework for app technology in the enterprise. *Procedia CIRP*, 7, 664-669. doi:10.1016/j.procir.2013.06.050
- Goldin, G., & Shteingold, N. (2001). Systems of representations and the development of mathematical concepts. *The roles of representation in school mathematics*, 1-23.
- Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning*, New York, NY: Springer.
- Kinnebrew, J. S., Segedy, J. R. & Biswas, G. (2014). Analyzing the temporal evolution of students' behaviors in open-ended learning environments. *Metacognition and Learning*, 9(2), 187-215
- Martin, T., & Schwartz, D. L. (2005). Physically distributed learning: Restructuring and reinterpreting physical environments in the development of fraction concepts. *Cognitive Science*, 29, 587-625.
- Moyer, P. S., Bolyard, J. J., & Spikell, M. A. (2002). What are virtual manipulatives? *Teaching Children Mathematics*, 8(6), 372-377.
- Moyer-Packenham, P. S., Anderson, K. L., Shumway, J. F., Tucker, S. I., Westenskow, A., Boyer-Thurgood, J. M., ... The Virtual Manipulatives Research Group at Utah State University. (2014). Developing research tools for young children's interactions with mathematics apps on the iPad. In *Proceedings of the 12th Annual Hawaii International Conference on Education (HICE)* (pp. 1685-1694). Honolulu, Hawaii.
- Moyer-Packenham, P. S., Bullock, E. P., Tucker, S. I., Shumway, J. F., Watts, C., Anderson-Pence, K. L., Westenskow, A., Boyer-Thurgood, J., Gulkilik, H., & Jordan, K. (2015). *The role of affordances in children's learning performance and efficiency when using virtual manipulatives mathematics iPad apps*. Unpublished manuscript.
- Moyer-Packenham, P. S., Shumway, J. F., Bullock, E., Tucker, S. I., Anderson-Pence, K. L., Westenskow, A., Boyer-Thurgood, J., Maahs-Fladung, C., Symanzik, J., Mahamane, S., MacDonald, B., Jordan, K. (2015). Young children's learning performance and efficiency when using virtual manipulatives mathematics iPad apps. *Journal of Computers in Mathematics and Science Teaching*, 34(1), 41-69.
- Moyer-Packenham, P. S., Westenskow, A., Shumway, J. F., Bullock, E., Tucker, S. I., Anderson-Pence, K. L., The Virtual Manipulatives Research Group at Utah State University. (2014). The effects of different virtual manipulatives for second graders' mathematics learning and efficiency in the touch-screen environment. Presented at the 12th International Conference of the Mathematics Education into the 21st Century Project, Herceg Novi, Montenegro.
- Moyer-Packenham, P. S., & Westenskow, A. (2013). Effects of virtual manipulatives on student achievement and mathematics learning. *International Journal of Virtual and Personal Learning Environments*, 4(3).
- Nemirovsky, R., Kelton, M. L., & Rhodehamel, B. (2013). Playing mathematical instruments: Emerging perceptuomotor integration with an interactive mathematics exhibit. *Journal for Research in Mathematics Education*, 44(2), 372-415.
- Paek, S. (2012). *The impact of multimodal virtual manipulatives on young children's mathematics learning* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Full Text. (UMI No. 3554708)

- Paek, S., Hoffman, D. L., & Black, J. B. (2013). Multi-modal interaction with virtual manipulatives: Supporting young children's math learning. In N. Rummel, M. Kapur, M. Nathan, & S. Puntambekar (Eds.), *10th International Computer Supported Collaborative Learning Conference* (Vol. 2, pp. 117–120). Madison, WI.
- Paek, S., Hoffman, D. L., Saravanos, A., Kim, M., & Black, J. B. (2012). Multi-modal interaction in digital instructional media. In C. Martin, A. Ochsner, & K. Squire (Eds.), *8th Annual Games, Learning, & Society Conference* (pp. 245–250). Madison, WI.
- Paek, S., Hoffman, D., Saravanos, A., Black, J., & Kinzer, C. (2011). The role of modality in virtual manipulative design. In D. Tan, B. Begole, & W. A. Kellogg (Eds.), (pp. 1747–1752). New York, NY: ACM. doi:10.1145/1979742.1979839
- Roschelle, J. (2000). Choosing and using video equipment for data collection. In A. E. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 709-731). Mahwah, NJ: Erlbaum.
- Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children*. New York: Routledge.
- Shuler, C. (2012). *iLearn II: An analysis of the education category of Apple's app store*. The Joan Ganz Cooney Center at Sesame Workshop.
- Tucker, S. I., & Moyer-Packenham, P. S. (2014). Virtual manipulatives' affordances influence student learning. In S. Oesterle, C. Nicol, P. Liljedahl, & D. Allan (Eds.), *Proceedings of the Joint Meeting of PME 38 and PME-NA 36* (Vol. 6, p. 251). Vancouver, Canada: PME.
- Tucker, S. I., Moyer-Packenham, P. S., Shumway, J. F., & Jordan, K. (2014). *Zooming in on students' thinking: How a number line app revealed, concealed, and developed students' number understanding*. Manuscript submitted for publication.
- Wilkinson, L., & Friendly, M. (2009). The history of the cluster heat map. *The American Statistician*, 63(2), 179–184.

Authors' Note: Financial support for the work reported in this article was provided by the VP for Research Office category of Research Catalyst Funding at Utah State University.

Appendix A: Full Names of Mathematics Apps Selected for Pre/Post Assessments and Learning Activities

App Name in Article	Full App Name*	App Section Used	App Developer*
100s Chart	100s Board	Full app	Matthew Thomas
Frog Number Line	Teaching Number Lines	Skip Counting Test	Aleesha Kondys
Counting Beads	Montessori Bead Skip Counting	Full app	MontessoriTech
Base-10 Blocks	Montessori Numbers—Math Activities for Kids	Quantity (100-999)	L'Escapadou
Zoom Number Line	Motion Math: Zoom	Levels 2-6	Motion Math
Place Value Cards	Montessori Place Value	3-digit numbers without zeros	MontessoriTech

* As listed on iTunes App Store