

Title:

Does “Measure Up!” measure up? Evaluation of an iPad app to teach preschoolers measurement concepts

Authors:

Katerina Schenke

Elizabeth J. K. H. Redman

Gregory K. W. K. Chung

Sandy M. Chang

Tianying Feng

Charles B. Parks

Jeremy Roberts

Journal publication date:

2020

Published in:

*Computers and Education*, Volume 146

Funded by:

U.S. Department of Education

Cooperative agreement number PRU295A150003

Does “Measure Up!” measure up? Evaluation of an iPad app to teach preschoolers measurement concepts

Katerina Schenke

Elizabeth J. K. H. Redman

Gregory K. W. K. Chung

National Center for Research on Evaluation, Standards, and Student Testing at the  
University of California, Los Angeles

Sandy M. Chang

Loyola Marymount University

Tianying Feng

Charles B. Parks

National Center for Research on Evaluation, Standards, and Student Testing at the  
University of California, Los Angeles

Jeremy Roberts

PBS KIDS Digital

**Acknowledgements.** We would like to thank Brettany Clemens, Claudia Riveroll, and Joanne Michiuye of CRESST for their help with data collection, as well as Lorrie Kato and her students at Los Angeles Harbor College. We would like to thank Alice Klein for providing consultation on our measures. Additionally, we would like to acknowledge the indispensable help extended to us by Rosa Serratore and Allyson Krogmann Jordan of the Santa Monica-Malibu Unified School District. Finally we would like to express our appreciation for the school sites, teachers, and students who made this research possible. The contents of this document were developed under a cooperative agreement #PRU295A150003, from the U.S. Department of Education. However, these contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government.

**Abstract**

Understanding digital supports for early learning is paramount for school readiness and later mathematics learning. We present results from a randomized control trial evaluating a digital app (Measure Up!) and a parent companion app (Super Vision) designed to teach children measurement concepts, a skill that many teachers do not feel comfortable teaching. Ninety-nine 4- and 5-year-old children were randomly assigned to one of three conditions: Measure Up!, Super Vision + Measure Up!, or a control game. Analyses revealed a statistically significant effect of being in the treatment group (Measure Up! or Measure Up! + Super Vision) on children's posttest scores (about two additional questions correct), controlling for the pretest and demographic characteristics (gender, SES). In particular, gains were made for children's understanding of pan balances. There was no significant difference between the two treatment groups. Results suggest that apps can be designed to help children learn important mathematics skills; however, more research needs to be done to understand how parent supports can be better designed. Implications for evaluation and design of game-based learning tools are discussed.

*Keywords.* Early childhood, digital games, mathematics, parental support, measurement knowledge

Does “Measure Up!” measure up? Evaluation of an iPad app to teach preschoolers measurement concepts

The allure of educational games stems from their ability to motivate, engage, and teach the children who play them. There is growing evidence that educational games can be effective for learning under certain conditions (Deater-Deckard, Mallah, Chang, Evans, & Norton, 2014; Ninaus, Kiili, McMullen, & Moeller, 2017; Schenke, Rutherford, & Farkas, 2014; Tobias, Fletcher, & Wind, 2014; for a meta analysis see Wouters & van Oostendorp, 2013), especially for young learners (see Aladé, Lauricella, Beaudoin-Ryan, & Wartella, 2016; Outhwaite, Faulder, Gulliford, & Pitchford, 2018; Räsänen, Salminen, Wilson, Aunio, & Dehaene, 2009; Schacter & Jo, 2016, 2017; Starkey, Klein, & Wakeley, 2004). In addition, there is growing evidence that well-designed games, when integrated into supportive home environments, can result in positive math outcomes for preschoolers (Pasnik, Moorthy, Llorente, Hupert, Dominguez, & Silander, 2015). For the purposes of this manuscript, we focus on the use of digital apps instead of games, which are commonly found for young learners in outlets such as Common Sense Media. One promising avenue for early intervention is the use of digital games and apps to target mathematical competencies and concepts. Digital games and apps may be an effective way to introduce preschoolers to academic concepts that they are missing, as children are becoming consumers of media at earlier and earlier ages (Common Sense Media, 2017). Digital games can provide positive learning experiences for children (Beschoner & Hutchinson, 2013; Lieberman, 2006). However, as promising as these technologies may appear, research in the area of digital games for mathematics learning has shown mixed results on their effectiveness for helping children learn (Dynarski et al., 2007; Ke, 2008; Kebritchi et al., 2010; Slavin & Lake, 2008; Suppes, Liang, Macken, & Flickinger, 2014). Additionally, whereas families are

increasingly using digital media with their young children (Common Sense Media, 2017), there is still relatively little research on educational digital media to teach preschool-aged children mathematics concepts (Schacter & Jo, 2016), with most research having been conducted around children's emerging literacy skills (Neumann & Neumann, 2014) and not on STEM content (see Aladé et al., 2016).

In this study, we sought to extend the research on using games and apps for learning into formal settings by examining how gameplay (and parent alerts related to their child's gameplay) affects children's learning of math concepts in preschool. We evaluated the effectiveness of the PBS KIDS Measure Up! app (MU)—developed by PBS KIDS for preschool and early elementary school learners—and a parent companion application to MU called PBS KIDS Super Vision (SV), which was developed for parents and caregivers to use to follow their children's progress (e.g., track the concepts and skills their children may have learned in MU). Our specific aim was to examine the effects of MU and SV on children's learning of measurement concepts and parents' awareness and support of their children's mathematics learning. We base our study design on a previous study conducted by WestEd (2015). The general research question investigated in this study was: To what extent did playing MU or playing MU in addition to using the parent-provided SV app impact children's learning of measurement concepts as compared with students who played a control app?

### **1.1 The Importance of Early Measurement Concepts**

The U.S. continues to lag behind many other developed nations in math achievement, as evidenced by the U.S.'s performance on international assessments such as the TIMSS and PISA (Fleischman, Hopstock, Pelczar, & Shelley, 2010; Provasnik, Malley, Stephens, Landeros, Perkins, & Tang, 2016). Additionally, when looking at mathematics performance across

countries, and even within the U.S., there are discrepancies in the academic achievement of students from different socioeconomic backgrounds. Specifically, children from socioeconomically disadvantaged families exhibit lower levels of math knowledge than their middle class peers (Aikens & Barbarin, 2008; Jordan, Kaplan, Locuniak, & Ramineni, 2007; Kozol, 1991; Oakes, 1990; Starkey & Klein, 2008; Starkey et al., 2004). These discrepancies between socioeconomic groups show up as early as preschool (Duncan & Magnuson, 2005; Lee & Burkam, 2002; Loeb & Bassok, 2008; Reardon & Robinson, 2008)—an important time in which early differences in mathematics knowledge develop and accumulate over time (Alexander, Entwisle, & Horsey, 1997; Brooks-Gunn & Duncan, 1997; Duncan, Brooks-Gunn, & Klebanov, 1994; Huston & Bentley, 2010). Because of the pre-existing differences in skills that occur before children even start school, there is a need for developing additional tools to help early learners.

The development of children's mathematics learning has long focused on areas around number sense, sequencing, and magnitude (see Clements & Sarama, 2008). However, measurement skills in early childhood have been recognized as an important area for improvement (see Battista, Clements, Arnoff, Battista, & Van Auken Borrow, 1998; Clements & Bright, 2003; Kamii & Kysh, 2006; Mullis, Martin, Gonzalez, & Chrostowski, 2004; National Center for Educational Statistics, 2018; Outhred & Mitchelmore, 2000) and are considered fundamental to math and science learning (Solomon, Vasilyeva, Huttenlocher, & Levine, 2015; Sophian, 2007). The domain of measurement in mathematics education focuses on concepts such as length, height, and weight. In general, the concept of measurement in early childhood refers to the assignment of a unit to a quantity such as inches to measure the length of a stick or non-standard units such as paperclips to measure the length of a stick. The development of

measurement skills such as understanding of mass, length, and weight begin in early childhood (Clements & Stephen, 2004). For example, young children around 4 to 5 years old can make perceptual judgments of which object is longer given two objects of different lengths (Clements & Stephen, 2004). Further, young children often use non-standard measurement units to measure such as using pennies to measure the length of a pencil (Cheeseman, McDonough, & Ferguson, 2014). Many tasks that children are given in preschool involve the use of blocks and paper strips to practice partitioning. Because measurement concepts are related to more advanced mathematical concepts such as fractions and decimals (Brendefur, Strother, Thiede, Lane, & Surges-Prokop, 2012; Cramer, Post, & Delmas, 2002; National Research Council 2001), they are fundamental to children's development of mathematical competency. As important as these mathematics skills are, measurement skills have been identified as an area of needed improvement as U.S. children score lower on topics of measurement and geometry than other topics in mathematics (see National Center for Education Statistics, 1996, 2009; Clements & Bright, 2003; Kamii & Housman, 2000; Mullis, et al., 2004)

Several misconceptions and difficulties about measurement concepts have been identified in the literature. For example, young children often have misconceptions about the use of pan balances and weight, not understanding that heavier objects push the weight of the pan balance down (Metz, 1993). Many children believe that the higher weight represents the heavier one (see Metz, 1993, p.48). Additionally, a study by Solomon and colleagues (2015) suggests that children have a hard time understanding that rulers represent a set of countable spatial unit intervals. Developing technologies that allow children to gain early practice on these skills could be fruitful to their development of measurement skills.

## **1.2 Using Digital Apps to Support Mathematics Learning**

Few studies have been explicitly focused on examining the impact of touch screens, including tablets, iPads, and smartphones, on young children (Crescenzi, Jewitt, & Price, 2014; Lieberman, Bates, & So, 2009; Neumann & Neumann, 2014; Outhwaite et al., 2017; Shacter & Jo, 2017; Starcic & Bagon, 2014). The claim of many of these digital interventions is that they support individualized learning or the self-pacing among children and that the use of multiple modalities can help children learn (see Aladé et al., 2016; Bosse, Jacobs, & Anderson, 2009; Gelman, Brenneman, Macdonald, & Roman, 2009; Pitchford, Kamchedzera, Hubber, & Chigeda, 2018). The benefit of interactivity for learning remains an open question. For example, Aladé and colleagues (2016) tested the interactivity of an app where they compared interactive condition, non interactive condition (where children watched a video), and a control condition. They did not find differences between the interactive and non interactive conditions on most of their measures (they had measures of near, medium, and far transfer). For the most difficult task (far transfer that was not scaffolded) the non interactive group did better. As such, technology that combines videos that are non interactive and games and manipulatives that are interactive might provide a plausible way forward in designing technology for young children.

### **1.3 Parent Support of Children's Mathematics Learning**

In addition to developing tools to help children with learning measurement constructs, tools that are targeted at parents could likely bolster children's learning of measurement concepts. Within the context of early childhood education, it has long been understood that parental involvement can have a positive effect on children's learning (Englund, Luckner, Whaley, & Egeland, 2004; Marcon, 1999; Miedel & Reynolds, 1999; see Jeynes, 2005 for a meta-analysis; Starkey & Klein, 2000). A meta-analysis by Jeynes (2005) of 41 studies reported an overall effect of parental involvement in the form of communication, help with homework,

reading, expectations, attendance, and specific parental involvement on academic achievement ranged from .70 to .75 of a standard deviation. Further, parental talk around mathematics is important in the development of a child's mathematics knowledge and this talk often occurs before children even enter school (Susperreguy & Davis-Kean, 2016). Specifically, parents engage in math talk most often around cardinal values and units of measure (Susperreguy & Davis-Kean, 2016), although the frequency with which parents engage in talk about cardinal values is more than double talk about units of measure. Moreover, though parents have been found to support their child's early math development, there are pronounced differences along the lines of socio economic status (SES; Levine, Suriyakham, Rowe, Huttenlocker, & Gunderson, 2010; Ramani, Rowe, Eason, & Leech., 2015) where higher SES families engage in different types of math talk than lower SES families (Levine et al., 2010). As such, addressing the achievement gap in preschool students' mathematics achievement may be accomplished by providing parents with supports that are directly linked to the child's learning.

#### **1.4 Present Study**

In the present study, we describe results from an evaluation of an iPad app called Measure Up! We ask the following two research questions:

- (1) Does playing Measure Up! (both with and without the parent companion app) impact children's mathematics learning compared with an active control group?
  - a. We hypothesize that children playing the Measure Up! app will experience higher gains in measurement knowledge compared to an active control group. We also hypothesize that children assigned to the Measure Up! + Super Vision condition will experience greater gains than children assigned to the

Measure Up! or control conditions due to the encouragement of the parental talk and activities with their children.

- (2) On which particular question categories (e.g., height, weight) did students in the MU and MU+SV conditions gain?
  - a. Because many of the games in the Measure Up! app contain opportunities for children to interact with pan balances, we hypothesize that children will make the most gains on the concepts involving pan balances.

To answer our research questions, we employ a randomized control trial where children were randomly assigned to one of three conditions: (1) Measure Up! app (MU-only), (2) Measure Up! app and the parent companion app called Super Vision (MU+SV), and (3) control games that is hypothesized to teach letter and word knowledge called Super WHY! (control).

## **Method**

### **2.1 Participants**

Participants were recruited between January and May 2018 from four school sites around a large city in the southwestern region of the United States. Three sites were public elementary schools, from which we recruited children from three prekindergarten classrooms and two transitional kindergarten classrooms. All three elementary schools were designated Title I schools. The fourth site was a childcare center affiliated with a community college; the median income in the surrounding neighborhood of the community college was reported to be \$42,112, in 2016 inflation-adjusted dollars (U.S. Census Bureau, n.d.). We recruited 4- and 5-year-olds from two classrooms from this childcare center. Once study sites agreed to participate, we recruited parents from those sites. Parents completed a screener survey, either taken online or given in person, to assess eligibility for their and their children's participation in the study.

Eligibility was determined if the child was 4 or 5 years old, and if the parent or guardian could read English (a requirement because SV is only available in English). After completing the screener, eligible parent-child pairs were randomly assigned within site to one of three conditions: Measure Up! only (MU-only), Measure Up! + Super Vision (MU+SV), and control. One hundred one participants were consented to participate in the study. One participant declined to complete the assessments and was not included in the final sample. Another participant was found to be ineligible (due to age) early on in the intervention and was excluded from further data collection. The final sample included 99 participants. Demographic information about the sample, including parental education, languages spoken in the home, is included in the online supplementary materials (Table S1). Forty five percent of our participants reported being Hispanic, 41% White, 10% African American, 11% Asian, and 7% reported being American Indian or other race/ethnicity. The majority of the sample (62%) reported a household income of \$50,000 or less. There were, however, a substantial number of families (22%) who reported their household income as over \$100,000. For our analyses, we included low-income dichotomous variable, which was derived from parents' reports of household income, household size and information from the Department of Housing and Urban Development for Los Angeles County for that year.

## **2.2 Procedure**

This study followed a pretest-posttest randomized design where participants were randomly assigned within site to be in one of three conditions: (a) MU-only where children played MU during class; (b) MU+SV where, in addition to children playing MU during class, parents were given SV and instructed to check the app periodically over the course of the intervention; and (c) a control condition where children played literacy games on iPads. The main difference

between the MU-only and MU+SV conditions is that parents in the MU+SV condition were asked to check SV several times during the study.

Children in the treatment conditions (MU-only, MU+SV) played MU four days a week for 20–30 minutes a day in their classrooms, for three weeks. Children in the control condition played Super WHY! games (literacy games) during those times. Whereas children were encouraged to play the entire session time, researchers did not require them to do so, and children were allowed to end the session early if they asked. Researchers did not scaffold children's gameplay. If a child asked a question while playing the game, the researcher responded with statements such as "I don't know, what do you think?", or "try to do that by yourself". Researchers did, however, answer questions about how to navigate through the app. For children who ended early, teachers allowed them to participate in choice time or other activities occurring in the classroom. Researchers were present for all gameplay sessions to monitor any technology issues.

Control condition children were separated from the children playing MU either by seating them on the floor or at tables away from those playing MU. MU-only and MU+SV children were intermixed and seated at desks in the classroom. The general procedure was designed to emphasize data quality: to ensure that children were using the correct iPad for their assigned condition and were using the appropriate save slot, monitor treatment fidelity, keep children in the control group separate from the treatment groups, and record any anomalies.

## **2.3 Intervention**

**2.3.1 Measure Up! app.** Children in the treatment conditions (MU-only and MU+SV) played the Measure Up! app, a multimedia app developed to teach preschool and early elementary school aged children measurement concepts such as height and length, weight, and

capacity. The app includes 11 games, 11 videos, five challenges (similar to assessments), and seven toys (open-ended activities that modeled cause and effect on measurement concepts, allowing players to interact with the concepts in a less structured way than games). PBS KIDS created challenges to engage players in activities that could act as in-app assessment items.

The app was divided into three worlds one for length/height (Treetop City), one for capacity (Magma Peak), and one for weight (Crystal Caves). Children in the treatment conditions (MU-only and MU+SV) were encouraged to explore Treetop City on the first day of gameplay, Magma Peak on the second day, and Crystal Caves on the third, with the rest of the intervention period for free play. However, this sequence was not enforced, only suggested, and children could interact with the app as they wanted. Short descriptions of the sequences for each world are described below.

The length and height games, videos, challenges and toys in Treetop City model the use of some basic vocabulary describing relative height and size (e.g., big, bigger, biggest). They also introduce the concepts of seriation, composition of length, nonstandard measurement, and the relationship between height and length. For example, one activity asks children to order mushrooms from shortest to tallest, or estimate the length of various crystals in terms of nonstandard units of measurement such as paper clips and dice.

The games, videos, challenges and toys in Magma Peak introduce children to the concept of capacity and displacement and teach basic vocabulary. Activities include estimating the amount of water needed to fill particular containers and experimenting with displacement by adding dinosaurs to a swimming hole to cause the water level to rise.

The games, videos, challenges and toys in Crystal Caves introduce players to the concept of weight, the use of pan balances, and related basic vocabulary (e.g., heavy, heavier). Activities

ask players to use pan balances to compare the weights of different items or order objects by their weight.

**2.3.2 Super Vision.** Parents of children in the MU+SV condition were given a phone loaded with SV, a companion app to MU. It allows parents to view their child's activity within MU and provides information about the concepts in the games and videos they are consuming, their proficiency scores for selected games, and activities for parents to do with their children to extend learning offline. Such activities included measuring items in the real world with nonstandard units, filling containers with water and conducting displacement experiments, and estimating relative weights of items by holding one in each hand. Parents were given instructions on how to access the app and were told that once the study started, they could access the app. Other than that, there was no additional instruction on how parents should participate or interact with the app. Parents could contact the research team at any point with questions.

## **2.4 Measures**

**2.4.1 Measurement knowledge assessment.** A 20-item assessment was administered to each child before and after the intervention. The assessment took approximately 10 minutes and was administered one-on-one with a researcher. In most cases a second researcher was present to observe and verify scoring. The 20 items were developed, adapted, or adopted to assess children's knowledge of (a) length and height; (b) capacity and displacement; and (c) weight. Four items were adapted from the Child Math Assessment (CMA; Starkey, Klein, & Wakeley, 2004) with slight format revisions, three items were adopted from the KeyMath-3 Diagnostic Assessment (Connolly, 2007), and 13 further items were developed by the researchers, including items to assess nonstandard measurement, use of a pan balance, and displacement. Of the 20 items, 11 employed manipulatives (objects children were asked to use to give their answer) to

pose questions, six used pictures, and four used an iPad app created by the researchers to simulate using a pan balance. We adopted the CMA's format and narrative to contextualize the assessment: participants were asked to help "Sparky" (a toy dog) with various tasks, such as selecting a long bone or weighing his toys on a pan balance. Nineteen of the 20 items were scored as either correct or incorrect with no possible partial credit. One item, however, asking about the function of pan balances, was scored on a 3-point scale (0, .5, 1) to capture nascent ideas about the use and function of pan balances. This item was scored by a team of two researchers and any disagreements were resolved and a consensus score applied. The reliability of the assessment as measured by alpha coefficient of internal consistency was .75 for the pretest and .82 for the posttest. The full assessment is available from the authors upon request.

**2.4.2 Parent background survey.** Parents filled out the parent background questionnaire at the start of the study. It asked parents about their relationship to the child, child's age in years and months, education, household income, household size, language use, child's ethnicity, and child's media use and familiarity with various PBS KIDS characters. Parents were asked to specify their household income according to six categories. Category one specified an income below \$25,000 a year; category two specified an income between \$25,001 and \$49,999 a year; category three specified an income between \$50,000 and \$74,999 a year; category four specified an income between \$75,000 and \$99,999 a year; category five specified an income above \$100,000; and category six allowed parents not to report household income. Of the 99 participants, 86 parents chose to disclose household income. We subsequently created an indicator for low-income status that combined information from parents' reported household income and the reported household size. Parents received a low-income indicator if they reported household income in one of the first two categories or had household size of four or more and

reported a total income in category three (between \$50,000 and \$74,999 a year; see U.S. Department of Housing and Urban Development, 2018 for a description of low-income status in Los Angeles County). Descriptive statistics for the sample by condition for the main variables of interest are found in Table 1.

## 2.5 Analytic Plan

**2.5.1 Ordinary least squares regression models.** To answer our first research question and evaluate whether students in the MU-only or MU+SV conditions performed better on the posttest than children who were assigned to the control condition, we conducted ordinary least squares (OLS) regression. The formula for the final model is given by:

$$Posttest_i = \beta_{0i} + \beta_1 Prettest_i + \beta_2 MeasureUp_i + \beta_3 SuperVision_i + \beta_4 Age_i + \beta_5 Sex_i + \beta_6 LowIncomeStatus_i + \lambda_7 SiteIndicator_i + e_i$$

where the score on the posttest for child  $i$  is determined by an overall intercept ( $\beta_{0i}$ ), their performance on the pretest ( $\beta_1 Prettest_i$ ), an indicator for condition ( $\beta_2 MeasureUp_i + \beta_3 SuperVision_i$ ), their age in months ( $\beta_4 Age_i$ ), their sex ( $\beta_5 Sex_i$ ), an indicator for low income status ( $\beta_6 LowIncomeStatus_i$ ), a vector of site-specific indicators such that each site was denoted as having a 0 or 1 value ( $\lambda_7 SiteIndicator_i$ ), and a child-specific error term ( $e_i$ ). To control for unobserved differences between sites, we included site-specific indicators such that treatment-control differences are evaluated within site (recall that children were randomly assigned within site).

## Results

**3.1 RQ1: Does playing Measure Up! or Measure Up! in addition to a parent using the Super Vision app, impact children's learning of measurement concepts compared with children who played a control app?**

Correlations among the variables of interest are in the supplemental materials (Table S2). We found a statistically significant correlation between how a student did on the pretest and how a student did on the posttest ( $r = .79, p < .0001$ ), low income status variable and pre and posttest scores ( $r = -.39, p < .0001$ , and  $r = -.41, p < .0001$ , respectively). Age in months was also positively related to scores on the pre- and posttests ( $r = .40, p < .0001$ , and  $r = .37, p < .0001$ , respectively).

In Table 2, we present results from four models. The coefficients in the table represent unstandardized results to aid in the interpretation of the results. Ideally, the inclusion of additional covariates in our model should not substantively affect our interpretation of the effect of the treatment conditions. We apply this same logic to the evaluation of the standard errors in our models. If, for example, we found that the inclusion of a specific variable substantively changed our estimated regression coefficient of the treatments (and also our standard errors), we would be skeptical of our results. Namely, our results should be robust to any additional covariates in our model. We subsequently report standardized effects as a measure of effect size (Cohen's  $d$ ) in the text of the results and not in the tables. The first model shows results of the intervention on posttest scores controlling only for pretest scores. The models increase in their complexity by adding covariates thus providing multiple models of robustness from which to build an argument for the validity of the results. We present these multiple models to assure that any third variables are not responsible for changes in the estimation of the treatment effects. The second model shows results of the intervention conditions only controlling for students' pretest score and site. To account for any unobserved differences between sites, we included site-level dummy variables. This decision was further justified because the randomization occurred within site. In Model 3, we additionally control for students' age (in months) and gender (boy or not).

Finally, in Model 4, we include a low-income status indicator. We note that the sample size differs in the last model where low-income status is added since 13 parents did not report income status on the demographic survey. Thus regression coefficients of Model 4 cannot be compared with any other regression coefficients across models. Only regression coefficients of models containing the same sample (i.e., sample size) can be compared. The regression coefficients for the treatment effects should be interpreted as raw scores. For example, a regression coefficient of 1.0 indicates that children in that group, on average, responded correctly to one additional question on the posttest as compared with children in the control condition.

Across all models, there was a high and statistically significant relationship between performance on the pretest and students' performance on the posttest (as expected). The  $R^2$ , a measure of the proportion of variance in the outcome variable that is explained by the variables in our model, ranged from .68 to .74 indicating that the predictors explain a substantial amount of variance in the outcome variable (students' scores on the mathematics posttest).

The first model suggests that controlling for a student's pretest score, students in the MU-only condition had statistically significantly higher posttest scores than students in the control condition ( $b = 2.26$ ,  $SE = .614$ ,  $p < .001$ , with a standardized effect size of  $B = .52$ ). Another way to think of it is that students in the MU-only condition, scored, on average, two more questions correctly on the posttest, or 11%. The effect of being in the MU+SV condition as compared with students in the control condition on a student's posttest score was also statistically significant ( $b = 1.646$ ,  $SE = .615$ ,  $p = .009$ , with a standardized effect size of  $B = .38$ ). This translates to an 8% gain on the posttest for the MU+SV condition.

Model 2 controls for student's pretest score and their site. Therefore, controlling for students' pretest scores and the site, the association between being in the MU-only condition (as

compared with students in the control condition) was statistically significant with an unstandardized regression coefficient of 2.07 ( $SE = .58$ ,  $p = .001$ , with a standardized effect size of  $B = .49$ ). This suggests that students in the MU-only condition, on average, gained 11% on the posttest (equivalent to just over two additional questions correct). The association between being in MU+SV as compared with the control condition was 1.59 and statistically significant ( $SE = .58$ ,  $p = .008$ , with a standardized effect size of  $B = .39$ ), suggesting that students in the MU+SV condition gained an average of 8% on the posttest.

Model 3 controls for students' age in months and whether the participant was a boy. As in the prior models, the magnitude of the unstandardized regression coefficients does not change with an unstandardized regression coefficient of 2.27 ( $SE = .59$ ,  $p < .001$ , with a standardized effect size of  $B = .53$ ). This suggests an 11% increase on the posttest for the MU-only condition. The effect of being in the MU+SV condition compared with the control condition was also statistically significant ( $b = 1.75$ ,  $SE = .59$ ,  $p = .004$ , with a standardized effect size of  $B = .43$ ). This suggests a 9% increase on the posttest for the MU+SV condition.

Finally, Model 4 presents results where we include an indicator of low-income status. Recall that the sample size changes from 99 students to 86 students (those whose parents reported income). The magnitude of the regression coefficient for the treatment effect of MU-only drops to  $b = 1.76$  ( $SE = .66$ ,  $p = .009$ , with a standardized effect size of  $B = .43$ ). This translates to a 9% gain for children in the MU-only condition. For the MU+SV condition, this regression coefficient drops to  $b = 1.61$  ( $SE = .65$ ,  $p = .016$ , with a standardized effect size of  $B = .41$ ). This translates to a 9% gain on the posttest for the MU+SV students.

Overall, we note that for models including the whole sample of students, there is a statistically significant, positive effect of MU-only on children's posttest scores as compared

with children in the control group and a statistically significant, positive effect of MU+SV children's posttests as compared with children in the control group. Additional significance tests showed that there were no statistically significant differences between the MU-only and MU+SV conditions.

### **3.2 RQ2: On which particular question categories did students in the MU and SV conditions gain?**

**3.2.1 Gains per question category.** The models created for the individual questions suffer from small sample sizes, low reliability, and thus a low credibility of the estimated effects. This can be alleviated by averaging the scores of the individual questions into one score per question category (see Table 3). The resulting score per question category ranges from 0 (none of the questions in this category was answered correctly) to 1 (all of the questions in this category were answered correctly). The same model formula is used as above with question category instead of individual questions. The coefficients can be interpreted similarly: a coefficient of .08 for the MU+SV group in the capacity category means that students in the MU+SV group answered 8% more capacity questions correct on the posttest than the control group. This equates to an additional  $(\text{number of questions in category}) * (\text{coefficient}) = 2 * .08 = 0.16$  correctly answered questions on the posttest. The effect of MU+SV on the weight questions on the iPad has a coefficient of .22, which indicates that an additional 22% of questions were answered correctly on the posttest by students in the MU+SV group in comparison to the control group. This equals  $4 * 0.22 = 0.88$  additional correctly answered questions on the posttest in comparison to the control group. While we do see small positive treatment effects across most of the question categories, the treatment effect is by far the largest for the weight questions delivered on the

iPad. The important finding from these models is that students in the treatment groups mostly improved in the weight questions that were asked on the iPad in comparison to the control group.

### **Discussion**

The purpose of this study was to provide evidence for the efficacy of an iPad app called Measure Up! and a parent-companion app called Super Vision designed to teach young children measurement concepts. We had three experimental conditions, one of which contained a supplemental app for parents that was linked to their child's gameplay in MU. Overall, we found that children who were randomly assigned to play the MU app made more gains on the posttest of measurement knowledge than children who were assigned to play a literacy app. Specifically, students answered, on average, two additional questions correctly on the posttest controlling for school site, age, and gender. However, despite theoretical support for the MU + SV condition, children in this condition did not make any additional gains over and above children in the MU app, suggesting that the parent companion app, in its current iteration, had no additional value above the MU app. When investigating where children made gains in their measurement knowledge, we found that they made the most gains on items that asked them to use a pan balance.

This study moves the field forward in providing evidence that informal learning environments and products can be designed to teach children concepts that they might otherwise not get in school or in areas where teachers may not feel adequately prepared to teach that content area. Although the context of the study was done in a controlled school setting (preschool and transitional kindergarten classrooms) the results of the study could be expanded to informal learning environments such as the home environment specifically because children were allowed to freely navigate the application. Additionally, whereas much of the previous

literature on digital mathematics interventions focus on one specific construct or a narrow range of constructs (see Solomon et al., 2015), we have provided evidence that multiple activities around the concept of measurement, such as the use of the pan balance and understanding of length and height, can provide opportunities for children to learn these concepts.

Part of this evaluation also concerned the use of a parent companion app called Super Vision. Whereas previous research would suggest that additional parental support about specific learning material (see the meta analysis from Jeynes, 2005) would provide an additional benefit over and above children just playing the app themselves, we did not find a meaningful difference between the performance of children assigned to the parent companion app condition and those who were not assigned to the parent companion app. In looking at data on parent's actual usage of Super Vision, we found that 30 of the 33 parents opened the SV app at some point during the study. Fewer parents, however, logged activities that were indicative of a substantive interaction with the app. Only five parents clicked on a link to take them to external resources that provided ways to engage their children in learning about measurement at home. Research on understanding how to better and more directly involve parents in their children's learning has been previously difficult for media companies. Evidence of early promise of technologies such as through text messages have been found promising (such as in Farrell, Smith, Reardon, & Obara, 2016; Kalil & Mayer, 2016) as opposed to asking parents to separately access an app.

#### **4.1 Limitations and Future Directions**

We outline several limitations of this work. First, we did not find any statistically significant differences between children in the MU-only condition and children in the MU+SV conditions. This is counter to what we would expect. The parent app was designed to augment the experience of the game for children. It was intended that parents engage in practices that

would supplement the content that his/her child was learning in the app during school. Therefore, we had hypothesized that children in the condition with the additional parent engagement app would do better than the MU-only condition, but this was not the case. However, that hypothesis was predicated on several assumptions: first, that parents would use the app to gain insight into their children's activity in Measure Up!, and second, that this would lead them to talk to their children about measurement concepts. It was not clear that these assumptions were founded. As part of the study we collected survey responses from parents about their self-reported usage of SV and also collected telemetry data from parents' actual usage of the app. Half of the sample answered affirmatively, that SV had encouraged them to talk to their children about their experiences with MU. However, when looking at the telemetry data from the Super Vision app, it was evident that there was quite a range of parental engagement with the app, with some parents registering as few as five clicks and some as many as 161. More specifically, 24 parents (73%) opened the activity list, which lists their child's activity in the MU app in general terms. Twenty-four (73%) opened activity details, which gives parents information on specific games played and videos watched, including descriptions of games and videos and their child's scores in those games (as available). However, only five (15%) clicked on links to take them to external resources that provided ways to engage their children in learning about measurement at home. Subsequent data collected from parent surveys also revealed that many parents also described technical challenges when trying to use Super Vision. It may be that if parents and children had played the app together, we might expect to see more parent engagement with the Super Vision app since parents would have a better understanding of what children were doing. We suggest that future research involving the understanding of parent engagement with technology collect

telemetry data to accurately represent parental engagement in the app as well as develop technology that is more user friendly.

Second, we found that children who received MU performed better on the items associated with the pan balance. Importantly, the assessment that the research team created for the purposes of this study combined items that involved tactile manipulatives *and* items administered on an iPad. Creating items administered on the iPad was purposefully done because the research team had difficulty finding a tactile version of a pan balance that could be reliably administered. Therefore, items measuring knowledge of pan balance are confounded with the medium with which they are administered. Therefore, it is unclear if participants performed better at completing tasks on the iPad or gained a better understanding of the pan balance and related concepts. However, since children in the control condition also played an iPad app and therefore could also have learned the basic controls of the iPad, we can rule out this explanation.

Third, the control group we used was asked to play a literacy app called Super Why! It is unclear how the results of the evaluation would change had we used a control group that played another mathematics app. It may be that the results that we see are a result of time on task (i.e., mathematics), and not really a result of the Measure Up! app. We suggest future evaluations of apps geared toward specific content areas be thoughtful in choosing what the control group is playing. An additional limitation is that the study was conducted in school-based setting. It would be interesting to see how the results of this evaluation would play out in the home setting. Per IRB regulations, we allowed children to stop playing the app when they wanted (and cut everyone off at 30 minutes). However, because this was done in a school setting with the children's classroom teacher, they may have felt pressured to continue playing the app longer.

Future direction could look at whether the same amount of time would be spent on the app had it been done in a totally informal setting (as was intended by the developers).

Though presenting an analysis of the telemetry data collected as part of this study is outside the scope of the current paper, we stress the importance of the ability to analyze the event-log data (i.e., telemetry data) associated with children's behaviors in the app. A future direction of this work will aim to better understand what specific games or tools were responsible for the learning gains and what dosage should be prescribed. The analysis of telemetry data and process data in general is potentially a powerful tool to detect the patterns of usage and patterns of play children exhibit (see Roberts, Chung, & Parks, 2016). Though this study only presents the results of the overall evaluation of the Measure Up! app and takes an "intent to treat" approach, we view that it is necessary to better understand children's game play patterns in a digital app. For example, our anecdotal evidence suggests that children were far more engaged in "toys" and "stickers" (things that did not have educational content) than they were in games that did poses educational content. Understanding these differential preferences among children could help designers of educational learning content better create environments that are engaging to learners.

### References

- Aikens, N. L., & Barbarin, O. (2008). Socioeconomic differences in reading trajectories: The contribution of family, neighborhood, and school contexts. *Journal of Educational Psychology, 100* (2), 235–251.
- Aladé, F., Lauricella, A. R., Beaudoin- Ryan, L., & Wartella, E. (2016). Measuring with Murray: Touchscreen technology and preschoolers' STEM learning. *Computers in Human Behavior, 62*, 433–441.
- Alexander, K. L., Entwisle, D. R., & Horsey, C. S. (1997). From first grade forward: Early foundations of high school dropout. *Sociology of Education, 87*–107.
- Battista, M. T., Clements, D. H., Arnoff, J., Battista, K., & Borrow, C. V. A. (1998). Students' spatial structuring of 2D arrays of squares. *Journal for Research in Mathematics Education, 503*–532.
- Beschorner, B. & Hutchison, A. (2013). iPads as a literacy teaching tool in early childhood. *International Journal of Education in Mathematics, Science and Technology, 1*(1), 16–24.
- Bosse, S., Jacobs, G., & Anderson, T. L. (2009). Science in the air. *YC Young Children, 64*(6), 10.
- Brendefur, J., Strother, S., Thiede, K., Lane, C., & Surges-Prokop, M. J. (2013). A professional development program to improve math skills among preschool children in Head Start. *Early Childhood Education Journal, 41*(3), 187–195.
- Brooks-Gunn, J., & Duncan, G. J. (1997). The effects of poverty on children. *The Future of Children, 55*–71.
- Cheeseman, J., McDonough, A., & Ferguson, S. (2014). Investigating young children's learning of mass measurement. *Mathematics Education Research Journal, 26*(2), 131–150.

- Clements, D. H., & Sarama, J. (2008). Curriculum focal points: Pre-K to Kindergarten. *Teaching Children Mathematics*, 14(6), 361–365.
- Clements, D. H., & Stephan, M. (2004). Measurement in pre-K to grade 2 mathematics. *Engaging young children in mathematics: Standards for Early Childhood Mathematics Education*, 299–317.
- Clements, D., & Bright, G. (Eds.) (2003). *Learning and teaching measurement, 2003 yearbook*. Reston, VA: National Council of Teachers of Mathematics.
- Common Sense Media. (2017). *The common sense census: Media use by kids age zero to eight*. Retrieved from [https://www.commonsensemedia.org/sites/default/files/uploads/research/csm\\_zerotoeight\\_fullreport\\_release\\_2.pdf](https://www.commonsensemedia.org/sites/default/files/uploads/research/csm_zerotoeight_fullreport_release_2.pdf)
- Connolly, A. J. (2007). *KeyMath 3: Diagnostic Assessment*. San Antonio, TX: Pearson.
- Cramer, K. A., Post, T. R., & Delmas, R. C. (2002). Initial fraction learning by fourth-and fifth-grade students: A comparison of the effects of using commercial curricula with the effects of using the rational number project curriculum. *Journal for Research in Mathematics Education*, 111–144.
- Crescenzi, L., Jewitt, C., & Price, S. (2014). The role of touch in preschool children's learning using iPad versus paper interaction. *Australian Journal of Language & Literacy*, 37(2), 86–95.
- Deater-Deckard, K., El Mallah, S., Chang, M., Evans, M. A., & Norton, A. (2014). Student behavioral engagement during mathematics educational video game instruction with 11–14 year olds. *International Journal of Child-Computer Interaction*, 2(3), 101–108.

- Duncan, G. J., & Magnuson, K. A. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps?. *The Future of Children*, 35–54.
- Duncan, G. J., Brooks-Gunn, J., & Klebanov, P. K. (1994). Economic deprivation and early childhood development. *Child Development*, 65(2), 296–318.
- Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L.,... Emery, D. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort*. Washington, DC: U.S. Department of Education.
- Englund, M. M., Luckner, A. E., Whaley, G. J., & Egeland, B. (2004). Children's achievement in early elementary school: Longitudinal effects of parental involvement, expectations, and quality of assistance. *Journal of Educational Psychology*, 96(4), 723–730.
- Farrell, M., Smith, J., Reardon, L., & Obara, E. (2016). *Framing the Message: Using Behavioral Economics to Engage TANF Recipients*. (OPRE Report 2016-02). Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Fleischman, H. L., Hopstock, P. J., Pelczar, M. P., & Shelley, B. E. (2010). Highlights from PISA 2009: Performance of US 15-year-old students in reading, mathematics, and science literacy in an international context. NCEES 2011-004. National Center for Education Statistics.
- Gelman, R., Brenneman, K., Macdonald, G., & Roman, M. (2009). *Preschool Pathways to Science (PrePS [TM]): Facilitating Scientific Ways of Thinking, Talking, Doing, and Understanding*. Baltimore, MD: Brookes Publishing Company.
- Huston, A. C., & Bentley, A. C. (2010). Human development in societal context. *Annual Review of Psychology*, 61, 411–437.

- Jeynes, W. H. (2005). A meta-analysis of the relation of parental involvement to urban elementary school student academic achievement. *Urban Education, 40*(3), 237–269.
- Kalil, A., & Mayer, S. E. (2016). Understanding the importance of parental time with children: Comment on Milkie, Nomaguchi, and Denny (2015). *Journal of Marriage and Family, 78*(1), 262–265.
- Kamii, C., & Housman, L. B. (2000). *Young children reinvent arithmetic* (2nd ed.). New York, NY: Teachers College Press.
- Kamii, C., & Kysh, J. (2006). The difficulty of “length× width”: Is a square the unit of measurement?. *The Journal of Mathematical Behavior, 25*(2), 105–115.
- Ke, F. (2008). A case study of computer gaming for math: engaged learning from gameplay? *Computers & Education, 51*(4), 1609–1620.
- Kebritchi, M., Hirumi, A., & Bai, H. (2010). The effects of modern mathematics computer games on mathematics achievement and class motivation. *Computers & Education, 55*(2), 427–443.
- Kozol, J. (1991). *Savage Inequalities: Children in America's Schools*. New York: Crown.
- Levine, S. C., Suriyakham, L. W., Rowe, M. L., Huttenlocher, J., & Gunderson, E. A. (2010). What counts in the development of young children's number knowledge?. *Developmental Psychology, 46*, 1309–1319.
- Lieberman, D. A. (2006). What can we learn from playing interactive games? In P. Vorderer & J. Bryant (Eds.) *Playing video games: Motives, Responses, and Consequences* (pp.379–397). Mahwah, NJ: Erlbaum.
- Lieberman, D. A., Bates, C. H., & So, J. (2009). Young children's learning with digital media. *Computers in the Schools, 26*(4), 271–283.

- Loeb, S. & Bassok, D. (2008). Early childhood and the achievement gap, In H.F. Ladd & E.B. Fiske (Eds.), *Handbook of Research in Education Policy Finance and Policy* (pp. 517–534). New York, NY: Routledge.
- Marcon, R. A. (1999). Positive relationships between parent school involvement and public school inner-city preschoolers' development and academic performance. *School Psychology Review, 28*(3), 395.
- Metz, K. E. (1993). Preschoolers' developing knowledge of the pan balance: From new representation to transformed problem solving. *Cognition and Instruction, 11*(1), 31–93.
- Miedel, W. T., & Reynolds, A. J. (1999). Parent involvement in early intervention for disadvantaged children: Does it matter?. *Journal of School Psychology, 37*(4), 379–402.
- Mullis, I. V., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Boston, MA: TIMSS & PIRLS International Study Center.
- National Center for Education Statistics. (1996). *Remedial education at higher education institutions in fall 1995 (Statistical Analysis Report)*, Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.
- National Center for Education Statistics, U.S. Department of Education. (2018). *2017 NAEP Mathematics and Reading Assessments: Highlighted Results at Grades 4 and 8 for the Nation, States, and Districts* (NCES Publication No. 2018037). Retrieved from <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2018037>
- National Research Council. (2001). *Preparing for an aging world: The case for cross-national research*. Washington, DC: National Academies Press (US).

- Neumann, M. M., & Neumann, D. L. (2014). Touch screen tablets and emergent literacy. *Early Childhood Education Journal*, 42(4), 231–239.
- Ninaus, M., Kiili, K., McMullen, J., & Moeller, K. (2017). Assessing fraction knowledge by a digital game. *Computers in Human Behavior*, 70, 197–206.
- Oakes J. 1990. *Multiplying inequalities: The effects of race, social class, and ability grouping on opportunities to learn mathematics and science*. Santa Monica, CA: RAND.
- Outhred, L. N., & Mitchelmore, M. C. (2000). Young children's intuitive understanding of rectangular area measurement. *Journal for Research in Mathematics Education*, 144–167.
- Outhwaite, L. A., Faulder, M., Gulliford, A., & Pitchford, N. J. (2018). Raising early achievement in math with interactive apps: A randomized control trial. *Journal of Educational Psychology*, 111(2), 284–298.
- Pasnik, S., Moorthy, S., Llorente, C., Hupert, N., Dominguez, X., & Silander, M. (2015). *Supporting parent–child experiences with PEG+CAT early math concepts: A report to the CPB-PBS Ready To Learn initiative*. Waltham, MA: Education Development Center; Menlo Park, CA: SRI International.
- Pitchford, N. J., Kamchedzera, E., Hubber, P. J., & Chigeda, A. L. (2018). Interactive apps promote learning of basic mathematics in children with special educational needs and disabilities. *Frontiers in Psychology*, 9, 262.
- Provasnik, S., Malley, L., Stephens, M., Landeros, K., Perkins, R., & Tang, J. H. (2016). Highlights from TIMSS and TIMSS Advanced 2015. *Mathematics Framework*, 1, 1–47.
- Ramani, G. B., Rowe, M. L., Eason, S. H., & Leech, K. A. (2015). Math talk during informal learning activities in Head Start families. *Cognitive Development*, 35, 15–33.

- Räsänen, P., Salminen, J., Wilson, A. J., Aunio, P., & Dehaene, S. (2009). Computer-assisted intervention for children with low numeracy skills. *Cognitive Development, 24*(4), 450–472.
- Reardon, S. F., Robinson, J. P., & Weathers, E. S. (2008). Patterns and trends in racial/ethnic and socioeconomic academic achievement gaps. In H. Ladd & E. Fiske (Eds.), *Handbook of research in education finance and policy* (pp. 497–516). New York, NY: Routledge.
- Roberts, J. D., Chung, G. K., & Parks, C. B. (2016). Supporting children's progress through the PBS KIDS learning analytics platform. *Journal of Children and Media, 10*(2), 257–266.
- Schacter, J., & Jo, B. (2016). Improving low-income preschoolers mathematics achievement with Math Shelf, a preschool tablet computer curriculum. *Computers in Human Behavior, 55*, 223–229.
- Schacter, J., & Jo, B. (2017). Improving preschoolers' mathematics achievement with tablets: A randomized controlled trial. *Mathematics Education Research Journal, 29*(3), 313–327.
- Schenke, K., Rutherford, T., & Farkas, G. (2014). Alignment of game design features and state mathematics standards: Do results reflect intentions? *Computers and Education, 76*, 215–224.
- Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: a best-evidence synthesis. *Review of Educational Research, 78*(3), 427–515.
- Solomon, T. L., Vasilyeva, M., Huttenlocher, J., & Levine, S. C. (2015). Minding the gap: Children's difficulty conceptualizing spatial intervals as linear measurement units. *Developmental Psychology, 51*(11), 1564–1573.

- Sophian, C. (2007, July). Measuring spatial factors in comparative judgments about large numerosities. In *International Conference on Foundations of Augmented Cognition* (pp. 157-165). Berlin, German: Springer.
- Starcic, A., & Bagon, S. (2014). ICT- supported learning for inclusion of people with special needs: Review of seven educational technology journals, 1970–2011. *British Journal of Educational Technology*, *45*(2), 202–230.
- Starkey, P., & Klein, A. (2000). Fostering parental support for children's mathematical development: An intervention with Head Start families. *Early Education and Development*, *11*, 659–680.
- Starkey, P. & Klein, A. (2008). Sociocultural influences on young children's mathematical knowledge". In O.N. Saracho & Spodek, B. (Eds.), *Contemporary perspectives on mathematics in early childhood education* (pp. 253–276). Charlotte, NC: Information Age.
- Starkey, P., Klein, A., & Wakeley, A. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Childhood Research Quarterly*, *19*(1), 99–120.
- Suppes, P., Liang, T., Macken, E. E., & Flickinger, D. P. (2014). Positive technological and negative pre-test-score effects in a four-year assessment of low socioeconomic status K–8 student learning in computer-based math and language arts courses. *Computers & Education*, *71*, 23–32.
- Susperreguy, M. I., & Davis-Kean, P. E. (2016). Maternal math talk in the home and math skills in preschool children. *Early Education and Development*, *27*(6), 841–857.

- Tobias, S. J., Fletcher, D., & Wind, A.P. (2014). Game-based learning. In J.M. Spector, M. D. Merrill, J. elen, & M.J. Bishop (Eds.) *Handbook of Research on Educational Communications and Technology* (pp.485–503). New York, NY: Springer New York.
- U.S. Census Bureau. (n.d.). *Selected economic characteristics: 2012-2016 American Community Survey 5-year estimates*. Retrieved from <http://factfinder.census.gov>
- U.S. Department of Housing and Urban Development. (2018). *FY 2018 income limits documentation system* [Data file]. Retrieved from <https://www.huduser.gov/portal/datasets/il/il2017/2017summary.odn>
- WestEd. (2015). *An early study of PBS KIDS Lost Lagoon and Super Vision apps. A report to the CPB-PBS Ready to Learn initiative*. San Francisco, CA: Author.
- Wouters, P., & van Oostendorp, H. (2013). A meta-analytic review of the role of instructional support in game-based learning. *Computers & Education*, 60(1), 412–425.

Table 1

*Descriptive Statistics for the Sample by Condition*

Variable	Total sample ( <i>N</i> = 99)				MU-only ( <i>n</i> = 33)				MU+SV ( <i>n</i> = 33)				Control ( <i>n</i> = 33)			
	<i>M</i>	<i>SD</i>	Min.	Max.	<i>M</i>	<i>SD</i>	Min.	Max.	<i>M</i>	<i>SD</i>	Min.	Max.	<i>M</i>	<i>SD</i>	Min.	Max.
Pretest score	10.38	3.71	1.0	19.0	9.62	3.57	1.0	15.5	11.18	3.65	2.0	18.5	10.35	3.85	2.0	19.0
Posttest score	12.00	4.32	2.0	20.0	12.24	4.50	2.0	19.5	13.09	3.63	4.0	19.0	10.67	4.54	3.0	20.0
Low-income indicator	0.71	—	0.0	1.0	0.70	—	0.0	1.0	0.76	—	0.0	1.0	0.67	—	0.0	1.0
Age in months	60.55	4.79	49.0	68.0	60.52	5.18	49.0	67.0	61.21	3.94	52.0	68.0	59.91	5.20	51.0	65.0
Participant was a boy indicator	0.52	—	0.0	1.0	0.64	—	0.0	1.0	0.55	—	0.0	1.0	0.36	—	0.0	1.0

*Note.* Raw scores on pre- and posttest are given. Indicator variables are dichotomous (0 or 1) and are reported as proportions

Table 2  
*Results from OLS Regression Predicting Posttest Mathematics Scores*

Variable	Model 1	Model 2	Model 3	Model 4
Pretest	0.933*** (0.069)	0.854*** (0.076)	0.873*** (0.077)	0.943*** (0.086)
MU-only	2.255*** (0.614)	2.072*** (0.581)	2.268*** (0.595)	1.763** (0.659)
MU+SV	1.646** (0.615)	1.587** (0.581)	1.752** (0.588)	1.606* (0.653)
Site 2		-1.797* (0.853)	-1.917* (0.853)	-1.687 (0.882)
Site 3		-0.214 (0.919)	0.084 (1.004)	0.148 (1.129)
Site 4		0.517 (1.008)	0.928 (1.069)	0.293 (1.258)
Site 5		-1.826* (0.909)	-1.611 (0.920)	-1.917* (0.957)
Age in months			-0.069 (0.067)	-0.072 (0.074)
Participant was a boy			-0.552 (0.500)	0.004 (0.569)
Low income status				-0.310 (0.763)
Constant	1.006 (0.832)	2.710** (0.963)	6.710 (3.828)	6.616 (4.180)
Observations	99	99	99	86
$R^2$	0.679	0.727	0.735	0.739

*Note.* Standard errors are in parentheses. The reference categories are control condition, girls, not low income, from Site 1.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

Table 3

*Coefficients for Hierarchical Linear Regression Models Predicting Question Categories on Posttest Score Based on Pretest Score, Treatment Condition, and Site*

Variable	Capacity (Q2-3)		Length (Q4-5, 8-10, 20)		Displacement (Q11)		Height (Q6-7, 18-19)		Weight (Q12-13)		Weight on iPad (Q14-17)		Weight paper and pencil (Q1-13, 18-20)	
	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>
Intercept	0.41	0.09	0.21	0.06	0.23	0.1	0.38	0.09	0.22	0.07	0.16	0.06	0.54	0.06
Pretest	0.43	0.09	0.61	0.09	0.46	0.09	0.58	0.11	0.6	0.09	0.53	0.1	0.2	0.1
MU+SV	0.08	0.08	0.06	0.05	0.16	0.11	0.03	0.06	0.05	0.07	0.22	0.07	0.05	0.07
MU	0.03	0.08	0.13	0.05	0.02	0.11	0.01	0.06	0.05	0.07	0.21	0.07	0.06	0.07

*Note.* The questions falling into the specific category are in parentheses. *B* are unstandardized regression coefficients and *SE* are standard errors.