A Study Comparing Virtual Manipulatives with Other Instructional Treatments in Third- and Fourth-Grade Classrooms

PATRICIA MOYER-PACKENHAM, JOSEPH BAKER, ARLA WESTENSKOW, KATIE ANDERSON, JESSICA SHUMWAY, KATI RODZON, AND KERRY JORDAN, THE VIRTUAL MANIPULATIVES RESEARCH GROUP AT UTAH STATE UNIVERSITY

ABSTRACT

The study reported here examined virtual manipulatives as an instructional treatment in 17 third- and fourth-grade classrooms. Students were randomly assigned to two treatment groups: texts and physical manipulatives (PM), and virtual manipulatives (VM). Results revealed no significant differences in achievement between the treatments. Additional results showed that objective ability predicted fraction achievement; virtual manipulative use can be modulated by test question type (e.g., symbolic vs. pictorial); percentage of class time using representations differed between VM and PM classrooms; and percentage of class time spent using representation types differed, potentially providing differential opportunities to learn.

INTRODUCTION

For the past 25 years, the use of technology (e.g., computers, iPads) has grown steadily in school mathematics classrooms. Technology has advanced from supporting simple programs to providing elaborate and sophisticated applications integrated with the Internet. Teachers regularly use virtual manipulatives for teaching mathematics in their classrooms using modalities that are mouse-driven for the PC, or that are manipulated on touch-screen devices. In some classrooms, the use of virtual manipulatives has replaced the use of physical manipulatives for mathematics instruction. Over the same period of time, educators and some parents have asked the question: Which is better—physical or virtual manipulatives? This is a complex question with complex answers.

Recently, two meta-analyses of research on manipulatives have been published. The first, conducted by Carbonneau, Marley, and Selig (2013), focused on the efficacy of teaching mathematics with physical (or concrete) manipulatives. The second, conducted by Moyer-Packenham & Westenskow (2012), focused on the effects of virtual manipulatives on student achievement. In their study, Carbonneau and colleagues identified 55 studies in which instruction with physical manipulatives was compared to instruction with abstract mathematical symbols. Results indicated that there were small to moderate effect sizes in favor of the use of the physical manipulatives and moderate to large effects of the manipulatives on retention. These results affirmed and extended the findings of Sowell (1989) who conducted the first meta-analysis on the effectiveness of physical manipulatives almost 25 years ago.

Moyer-Packenham and Westenskow (2012) identified 32 studies with 82 effect size scores for their meta-analysis, which examined research on instruction with virtual manipulatives compared to a variety of instructional treatments (e.g., instruction with physical manipulatives, instruction with abstract mathematics symbols, instruction with a combination of both physical and virtual manipulatives). Similar to Carbonneau et al. (2013), Moyer-Packenham and Westenskow found small to moderate effect sizes in favor of the use of the virtual manipulatives when compared with other instructional treatments. These meta-analyses provide support for some of the claims of the effectiveness of manipulative use for mathematics instruction.

While the meta-analyses conducted by Carbonneau et al. (2013) and Moyer-Packenham and Westenskow (2012) provide some insights on this question, they also reveal gaps in the research literature on physical and virtual manipulatives.

One gap is a lack of reported studies comparing physical and virtual manipulatives that include multiple elements of rigor in the methodology and research design. Previous studies comparing instruction using physical and virtual manipulatives have included some rigorous design elements, but have not included multiple elements of a rigorous experimental or quasi-experimental design. For example, only eight studies in the meta-analyses described above included random assignment of students to treatment groups when comparing physical and virtual manipulatives; only four studies included a large sample size (i.e., over 200 participants); only two studies examined the delayed effects of the treatment (i.e., delayed post testing); no studies included clear measures to assess the instructional fidelity of the treatments; and, no studies reported the psychometric properties of the instruments used. To truly understand the effects of physical and virtual manipulatives as instructional treatments in mathematics, a study that includes all these elements of a rigorous design is needed.

The purpose of the research that is the focus of this article was to address the need for a current and rigorous design in the study of physical and virtual manipulatives as instructional treatments in mathematics by incorporating multiple elements that have been absent, or not combined, in prior research. For this study, we adopted Moyer, Bolyard, and Spikell's (2002) definition of a virtual manipulative: "an interactive, web-based visual representation of a dynamic object that presents opportunities for constructing mathematical knowledge" (p. 373).

This study of physical and virtual manipulatives as instructional treatments in mathematics in 17 third- and fourth-grade classrooms randomly assigned students to the two treatments during

fraction instruction, developed instruments to assess student learning using Item Response Theory (IRT), observed instruction to determine treatment fidelity, and assessed students on measures of learning (i.e., post-tests) and retention (i.e., delayed post-tests). Including all these important research design elements in a single study contributes important insights on the effects of physical and virtual manipulatives as instructional treatments in mathematics on student achievement.

RESEARCH QUESTIONS

The overarching research question was: Are there differences in achievement on fraction learning and retention between third- and fourth-grade classrooms using virtual manipulatives fraction applets in a computer lab (VM) and those using texts and physical (concrete) manipulatives in a regular classroom (PM), as indicated by scores on pre-tests, post-tests, and delayed post-tests? Within the context of this broad question, we addressed the following sub-questions: a) Does student *objective ability,* based on pre-test scores in relation to the class pre-test average, predict fraction achievement, learning, and retention in either VM or PM classrooms? b) Is any impact of virtual manipulative use on students' fraction achievement modulated by mathematics *content test question type* (e.g., symbolic, pictorial, and combined)? c) Does percentage of class time spent in different *instructional configurations* (e.g., *groups* vs. *individually*) differ between VM and PM classrooms? d) Does percentage of class time spent using each type of fraction *representation* (e.g., pictorial, symbolic, manipulative type) differ between VM and PM classrooms? and e) Are there interaction effects of class time spent in different instructional configurations by fraction representation?

REVIEW OF THE LITERATURE

The Foundations of the Research on Manipulatives

The use of manipulatives (first physical manipulatives, now virtual manipulatives) has a long historical trajectory leading to their prevalence and use in mathematics classrooms today. The research has a 40-year history (Fuson & Briars, 1990; Moyer, 2001; Moyer & Jones, 2004; Parham, 1983; Prigge, 1978; Raphael & Wahlstrom, 1989; Sowell, 1989; Suydam, 1985; Suydam & Higgins, 1977; Thompson, 1992; Uribe-Florez & Wilkins, 2010). Studies of virtual manipulatives began over 25 years ago with the first computer-based manipulatives (Berlin & White, 1986; Clements & Battista, 1989; Clements & Sarama, 2007; Moreno & Mayer, 1999; Moyer-Packenham & Westenskow, 2012; Reimer & Moyer, 2005; Thompson, 1985).

Foundational theories on mathematics learning have led to the prevalence of manipulatives (both physical and virtual) for mathematics instruction. Over 50 years ago, Piaget's (1952) findings of clinical interviews suggested that children need experiences with the physical manipulation of objects to support their learning of abstract mathematical ideas. Bruner (1960, 1986) proposed that students' understanding occurred in three stages, the first stage an enactive stage where students interact with objects (i.e., mathematics manipulatives), prior to the iconic and symbolic stages. Zoltan Dienes (1969) suggested that students need multiple embodiments of a concept, and Dienes Blocks (a physical manipulative set of blocks) were developed for students to manipulate during mathematics experiences to promote learning.

Theories of cognition and the social construction of knowledge (Cobb, 1995; Vygotsky, 1978) consider manipulatives to be cognitive and cultural tools that are negotiated in the teaching and learning relationship. However, researchers have argued that the mathematics does not reside in the blocks themselves, and that manipulative use can be problematic for teaching and learning. For example, Ball (1992) argues, "Although kinesthetic experience can enhance perception and thinking, understanding does not travel through the fingertips and up the arm" (p. 47). Additionally, Meira's (1998) research suggested a caution to teachers that manipulatives, which are only the manufacturer's representation of a mathematical concept, have different degrees of transparency. Meira defined transparency as "an index of access to knowledge and activities rather than as an inherent feature of objects…a process mediated by unfolding activities and users' participation in ongoing sociocultural practices" (p. 121). Therefore, the physicality of the objects does not carry mathematical meaning. Meaning can only be constructed when students reflect on their actions with the manipulatives (whether physical or virtual). Additionally, recent studies on the use of manipulatives by K–8 teachers show that grade level and teacher beliefs and experience with the manipulatives are important predictors of how effectively teachers use them with students during mathematics instruction (Moyer-Packenham, Salkind, Bolyard, & Suh, 2013; Uribe-Florez & Wilkins, 2010).

In the recently adopted *Common Core State Standards for Mathematics* (CCSSM, National Governors Association Center for Best Practices—NGACBP & Council of Chief State School Officers—CCSSO, 2010) eight Mathematical Practices are described, including the expectation that students use appropriate tools strategically. Tool use includes both physical and virtual tools (i.e., physical and virtual manipulatives). Constructivist theorists suggest that learning is mediated by tools and therefore, the "tool changes the form, structure, and character of the activity" (Duffy & Cunningham, 1996, p. 19). Hiebert et al. (1997) suggest that ". . . different tools are different forms of representation, and each conveys a somewhat different message, and each emphasizes somewhat different features of the idea" (p. 58). If the CCSSM require students to "use appropriate tools" and to use those tools "strategically," it will be important for teachers to understand how students select tools for their own use when given the opportunity (Moyer & Jones, 2004) and how students employ tools to strategically solve problems (Schoenfeld, 1983).

Research using microgenetic analysis of students' problem solving has revealed that students who have a deep understanding of the relationships among different representations are able to use this knowledge in unfamiliar mathematical situations, while those

without this knowledge resist a change in their conceptual structures because their misconceptions are deeply rooted and robust (Schoenfeld, Smith, & Arcavi, 1993). Although early research using microgenetic analysis to examine students' translation among representations (e.g., graphical, symbolic, and abstract) began with students using paper and pencil, technology tools (like virtual manipulatives) can also be a source for examining students' translation among representations. As Lesh and Doerr (2003) suggest, ". . . these new conceptual tools are more than simply new ways to carry out old procedures; they are radically expanding the kind of problem solving and decision-making situations that should be emphasized in instruction and assessment" (p. 15). Technology tools like virtual manipulatives allow students to visualize, experiment, observe, reorganize, design, construct, and obtain feedback, and these actions extend students' experiences with representations beyond paper and pencil (Arcavi & Hadas, 2000). Both representational fluency (Zbiek, Heid, Blume, & Dick, 2007) and representational systems (Goldin, 2003) are important in the construction of mathematical meaning and sense making.

Design Methods Used in Previous Research on Manipulatives

Our review identified 21 peer-reviewed articles and 11 dissertations/theses that used quantitative methods to compare the effects on student achievement when virtual manipulatives were compared with other instructional treatments. As reported in the meta-analysis by Moyer-Packenham and Westenskow (2012), and of interest to the present study, were effects that focused on fraction instruction in third and fourth grade. Moyer-Packenham and Westenskow reported moderate effects for virtual manipulatives when compared with other instructional treatments in studies of fraction instruction ($f = 0.53$) and those conducted in third and fourth grade ($f = 0.37$). Also of interest were five elements that we considered to be important to a rigorous design: 1) random assignment of students to treatment groups, 2) large numbers of participants (i.e., > 200), 3) delayed post-testing to determine long-term retention effects, 4) assessment of the instructional fidelity of the treatments, and 5) the use of IRT to develop assessments specific to the study.

Among the 32 studies, many employed random assignment of intact classrooms to treatment groups, but only 10 studies randomly assigned individual students to treatment groups (Berlin & White, 1986; Burns & Hamm, 2011; Clements & Battista, 1989; Dinardi-Besterman, 1992; Hauptman, 2010; Martin & Lukong, 2005; Mendiburo & Hasselbring, 2011; Nute, 1997; Smith, 2006; Steen, Brooks, & Lyon, 2006). Of these ten studies with true random assignment, only one included a large number of participants over 200 students (Nute, 1997). Small numbers of participants could have affected the results as a confound given the differences in the pre-treatment ability levels of the groups.

Four studies had more than 200 participants (Clements, Battista, & Sarama, 2001; Nute, 1997; Pleet, 1991; Terry, 1995). Of these, three were dissertations, and there was no evidence that

they were then published in peer-reviewed journals. The study with the largest number of participants (1,055) examined the use of Logo Geometry software (Clements, Battista, & Sarama, 2001). The other studies ranged from 241 to 560 participants. All four studies were conducted over 10 years ago and used either researcher-designed applets or mathematics software, making these technologies different from current virtual manipulatives, which contain many unique affordances for learners.

Our review located only two studies that employed delayed post-testing to determine long-term retention effects of virtual manipulatives as a treatment (Clements, et al., 2001; Lin, 2010). Clements et al. reported that scores of the group using Logo geometry software increased significantly from the previous test to the delayed test, while the scores of students in traditional instruction decreased. Results of Lin's study showed that pre-service teachers using virtual manipulatives outperformed pre-service teachers using traditional instruction in procedural ($f = 0.15$) and conceptual ($f = 0.17$) fraction knowledge, with a significant difference on pre, post, and delayed tests following a four-week delay.

We located only six studies in which virtual manipulatives were used for instructional comparisons of fraction concept learning in elementary classrooms that included pre- and post-testing to determine statistically significant differences among the treatment groups and effect sizes (Ball, 1988; Burns & Hamm, 2011; Melideo & Dodson, 2009; Mendiburo & Hasselbring, 2011; Moyer-Packenham & Suh, 2012; Suh & Moyer-Packenham, 2007). For example, in examining the instruction of fraction addition algorithms, Ball compared the use of virtual manipulatives with traditional instructional techniques in five classes of fourth-grade students. Suh and Moyer-Packenham compared the use of physical and virtual manipulatives in the instruction of fraction addition to 36 third-grade students. In both studies, the classes that used virtual manipulatives significantly outperformed the groups that did not. However, the other four studies did not report statistically significant differences between treatment groups, although there were pre- to post-test gains for all groups. For example, Melideo and Dodson found no significant differences between groups when comparing physical and virtual manipulatives during fraction instruction with 20 fourth-grade students during a 9-day unit, and Mendiburo and Hasselbring found no significant differences between groups of fifth graders in a 10-day unit on fractions. Moyer-Packenham and Suh conducted the same type of comparison (i.e., physical vs. virtual) with 24 fifth-grade students spending 5 days using virtual manipulatives and also found no significant differences between groups. Burns & Hamm's study of 156 third- and fourth-grade students also revealed no statistically significant differences, but the treatment only lasted for one class session (i.e., 60 minutes).

The summary of research on physical and virtual manipulatives reveals a variety of limitations in research methods. Because of these limitations, confounding factors may account for the positive, negative, or neutral results that have been obtained in these studies. Our review points to the need for studies that compare

physical and virtual manipulatives and that include multiple elements of a rigorous design. The present study addresses this need by including physical and virtual manipulatives as instructional treatments, a large number of participants, random assignment of students to treatments, rigorously developed mathematics assessments, measures that assess the instructional fidelity of the treatments, and delayed post-tests to assess retention. This study makes an important contribution to the literature by combining multiple rigorous research design elements in a single study that examines physical and virtual manipulatives.

METHODOLOGY

### Participants

Three hundred fifty students (Third grade, N = 156; Fourth grade, N = 194) from 17 public school classrooms, located in 7 different elementary schools in 3 school districts, participated in the study. Researchers randomly assigned participants within-class to one of two treatment groups. One of the treatment groups used texts and physical manipulatives in regular classrooms (PM); the other treatment group used virtual manipulatives in a computer lab (VM).

Texts and physical manipulatives (PM) treatment group. The 17 PM classes were taught by 14 different third- and fourth-grade public school classroom teachers. The teachers' experience ranged from 3–32 years, with a mean of 17.6 years; all except two had taught third or fourth grade for three years or more. In a survey administered to the teachers, 64.3% reported that their students used *technology* every day prior to this study, while 37.7% reported that their students used technology once or twice a week. The teachers defined technology use as the SMART Board™, computers, and math programs. When asked about students' use of *computers*, 39.3% of teachers reported that their students used computers every day, while 60.7% reported their students used computers at least once or twice a week. Half of the teachers had never used the *National Library of Virtual Manipulatives (NLVM)* with their students, 14% reported using the NLVM website about twice per month, and 36% reported using it a few times a year.

During the treatment, all the classrooms used Pearson Success-Net curriculum materials (Scott Foresman/Addison Wesley Mathematics 2005 textbook). The instructional worksheets for students provided by the publisher include pictorial models and numeric and text-based information on each worksheet. (See Figure 1.) These materials were used by the instructors in the PM groups during the study. Twelve of the 17 classrooms (70.5%) also used SMART Board™ technology during mathematics instruction to display some of these images for students during the lessons. As can be seen from the worksheet in Figure 1, the pictorial models on the student worksheet match the physical manipulatives available for students that are shown in Figure 2. When teachers used the physical manipulatives during mathematics instruction, they demonstrated what they wanted students to do with the tools in order to solve mathematics problems. In these lessons, teachers

Figure 1. Example of the Pearson SuccessNet Curriculum Materials in the Mathematics Textbook Published by Scott Foresman/Addison Wesley (2005).
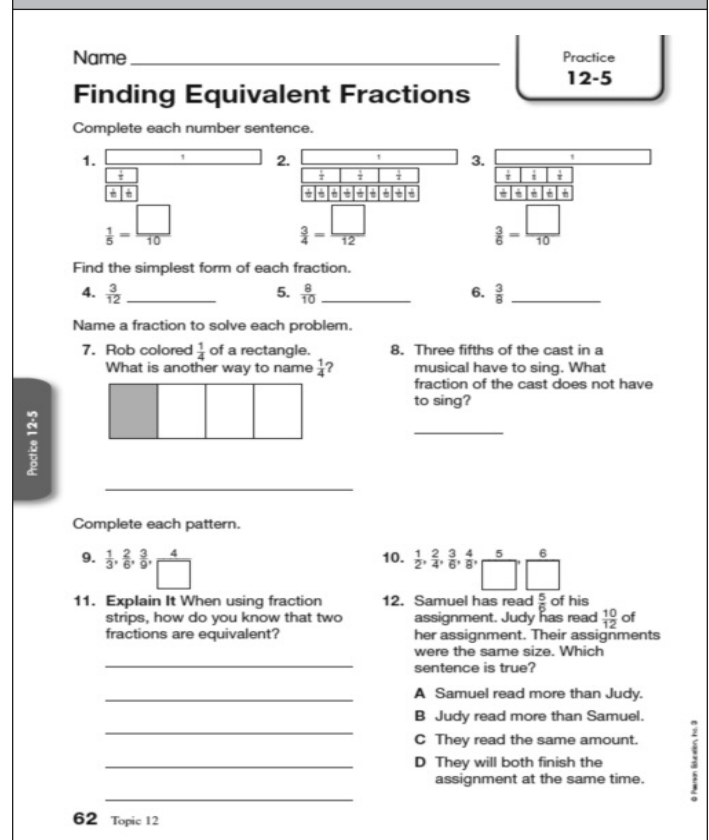
introduced a concept using the manipulatives. Then teachers allowed students to use the physical manipulatives (e.g., fraction pies or fraction tiles) while completing the worksheet.

The most common physical manipulatives used during fraction instruction included fraction pies and fraction tiles (see Figure 2). The fraction pies are a circular region representation of the whole divided into different numbers of fractional parts. The fraction tiles are a length model where one whole length is divided into different numbers of fractional parts. Fraction pies and fraction tiles can be used for comparison and to find equivalent portions.

Teachers also used other fraction manipulatives, worksheets, and teacher-created resources (e.g., paper cut-outs of fraction
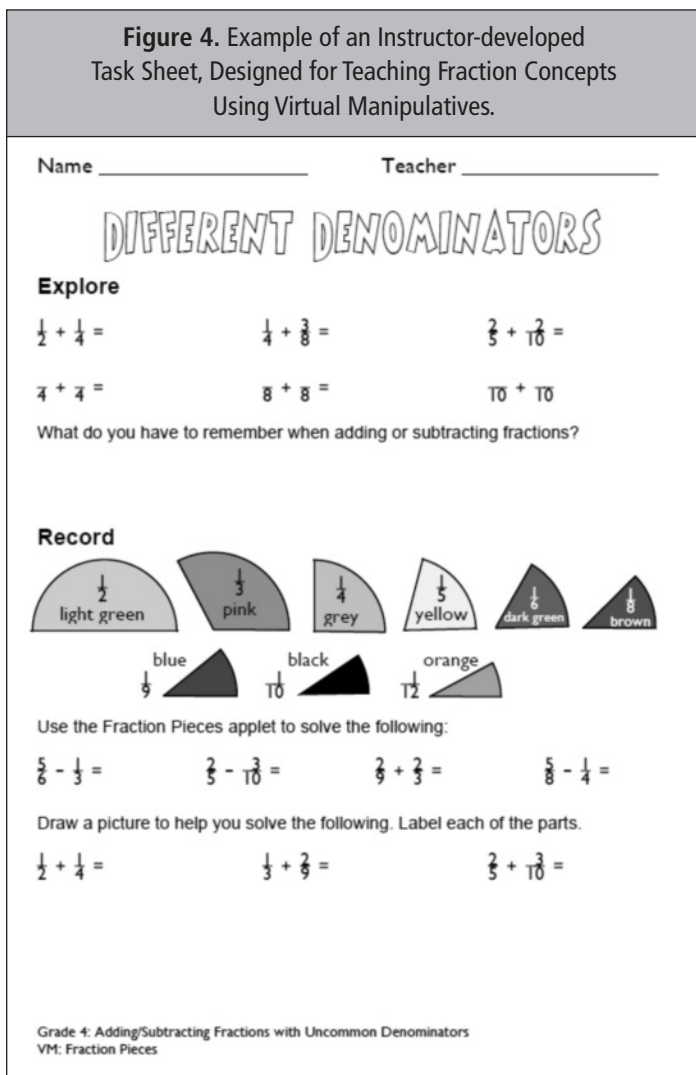

Figure 2. Physical Manipulatives (Fraction Pies and Fraction Tiles) Used in the PM Treatment Groups.

Figure 3. Virtual Manipulatives (Fraction Pies and Fraction Tiles) Used in the VM Treatment Groups.
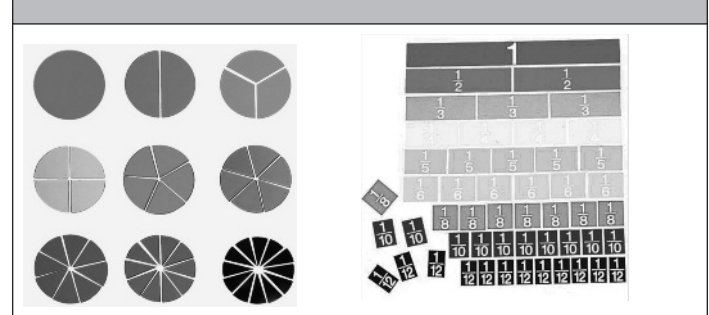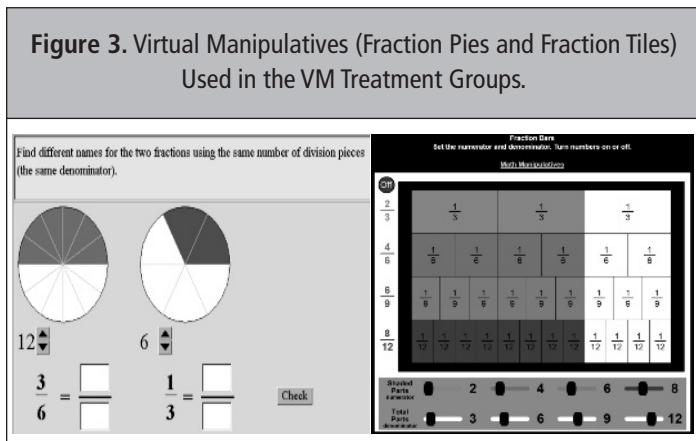
regions) during the fraction unit. Over 43% of teachers reported that they did not use computers or computer programs during the study, 28% reported using computers—but not virtual manipulatives—and 28% used computers every day, but again, not virtual manipulatives.

Virtual manipulatives (VM) treatment group. Four individuals affiliated with the local university taught the VM groups, including three doctoral-level graduate students and one university faculty mem-

Figure 4. Example of an Instructor-developed Task Sheet, Designed for Teaching Fraction Concepts Using Virtual Manipulatives.

ber (all former elementary teachers). The university teachers' public school teaching experience ranged from 7–30 years, with a mean of 14.6 years, and three of the four had public school experience teaching third or fourth grade.

The students in the VM treatment groups used virtual manipulative fraction applets from a variety of websites. Examples of the virtual manipulative applets included virtual fraction pies and virtual fraction tiles (see Figure 3). The virtual fraction pies are a representation of a circular region and the "arrow keys" on the virtual tool allow students to divide the circular region into different numbers of fractional parts. The virtual fraction tiles are a representation of a length model. Students can use a "slider" on the virtual tool, which allows them to create different numbers of fractional parts and shade the fractional parts for comparison. On both the fraction pies and the fraction tiles, numeric information accompanies the visual models that are in the virtual manipulatives tools.

During each lesson in the VM treatment groups, the instructors began with an introduction to the mathematics concept and to the virtual manipulatives that would be used by meeting with the whole group of students. Instructors demonstrated the keys and how to navigate within each of the virtual manipulative tools. Next the students interacted with one or more of the virtual manipulative applets independently to complete mathematics tasks. These independent interactions and explorations were guided by a task sheet that was specifically designed to teach fraction concepts using virtual manipulatives. An example of one of the VM task sheets is shown in Figure 4.

During the VM lessons, the instructors moved about the computer classroom and interacted with individual students to provide guidance and feedback and facilitate students using the virtual manipulatives to complete the guided task sheets. Students worked at their individual computers and task sheets at their own pace. At the end of each lesson, teachers pulled students back together as a whole group for a summary discussion of the day's concept.

The VM instructors also used some of the Pearson SuccessNet curriculum materials that were used by the PM instructors in addition to the VM instructor-developed task sheets. Tasks specific to problem exploration using the virtual manipulatives enabled VM task sheets to mirror the mathematical content being taught to the PM group. An expert group of experienced teachers had evaluated the lesson materials to determine the mathematical content match between PM and VM lessons. In preparation for the research project, the lesson materials had been piloted in test classrooms, reviewed, and revised, as necessary.

Procedures

Student demographics of gender, race, English Language Learner (ELL) status; socio-economic status (SES); and two measures of mathematical ability, subjective and objective, were reported by each classroom teacher at the beginning of the study. *Subjective* mathematical ability was determined by teacher rating of students' mathematical ability as high, medium, or low, based on their

knowledge of students' prior mathematics performance. The research team rated students' *objective* mathematical ability by comparing each student's pre-test score to the class pre-test average and standard deviation. Standardized scores one standard deviation or more below the mean were classified as low, while scores one standard deviation or more above the mean were classified as high. The remaining scores were rated as medium. At the end of the unit on fractions, classroom teachers identified any student absent for more than 40% of the time. Subsequent data analyses did not include information about these students. Teachers reported additional data concerning factors possibly influencing the instructional environment.

The design of the study ensured instructional fidelity across PM and VM treatment groups. Before beginning instruction, each paired teacher met to specify the number of days allotted for the fraction unit and to correlate lessons with the state's mathematics curriculum. This collaboration ensured that students received instruction on the same mathematical content regardless of treatment group. In an effort to address possible teacher effects, the instructors in the PM and VM treatment groups all had a minimum of three years of teaching experience, and the mean number of years of teaching experience was similar for the two treatment groups (17.6 years for the PM group and 14.6 for the VM group). Additionally, both groups of teachers had experience teaching elementary school, with almost all of the instructors having taught in Grades 3 and 4 previously. Over 70% of the lessons were observed to ensure that there were no differences in the mathematics content that students learned during the lessons, and that the instructional materials and strategies were documented for analysis. To further ensure conformity in lesson plans between treatment groups, each set of paired teachers met after each day's lesson to discuss plans for the next day. If the teachers decided that students were struggling with a particular concept, the pair of instructors together decided to re-teach that concept. Thus, daily check-ins ensured that students in both treatment groups learned the same content each day. Finally, null statistical comparisons of learning and retention outcomes between PM and VM instruction groups further reinforces our claim that teacher effects were not likely to contribute to the significant effects reported here.

Treatment groups met daily, and all fraction instruction occurred during regularly scheduled mathematics classes. VM treatment groups spent the fraction unit in the computer lab, using individual computers for approximately 50 minutes each day. Excluding administration of pre- and post-tests to both treatment groups by the classroom teachers, the duration of the fraction unit in each of the 17 classrooms ranged from 9 to 17 days (avg. = 11 days).

Third-grade lessons addressed the following concepts: identify the denominator of a fraction as the number of equal parts of the unit whole and the numerator of a fraction as the number of equal parts being considered; define regions and sets of objects as a whole, and divide the whole into equal parts using a variety of objects, models, and illustrations; name and write a fraction to represent a portion of a unit whole for halves, thirds, fourths, sixths, and eighths; place fractions on the number line, and compare and order fractions using models, pictures, number line, and symbols; and find equivalent fractions using concrete and pictorial representations.

Fourth-grade lessons addressed the following concepts: divide regions, lengths, and sets of objects into equal parts using a variety of models and illustrations; name and write a fraction to represent a portion of a unit whole length or set for halves, thirds, fourths, fifths, sixths, eighths, and tenths; generate equivalent fractions, and simplify fractions using models, pictures, and symbols; order simple fractions; use models to add and subtract simple fractions where one single digit denominator is one, two, or three times the other; add and subtract simple fractions where one single digit denominator is one, two, or three times the other.

Instruments

Mathematics content tests and observation ethograms were used to collect data (MacNulty, Mech, & Smith, 2007). Three mathematics content tests were administered: a pre-test immediately prior to the fraction unit, a post-test the day after the conclusion of the fraction unit, and a delayed post-test administered six to eight weeks after the fraction unit concluded. Throughout the study, observation ethograms documented instruction in each classroom. The following sections describe the development and administration of these instruments.

Pre- and post-tests. For the fractions pre- and post-tests, items were taken and/or adapted from four standardized test databases: National Assessment of Educational Progress (NAEP, 2011), Massachusetts Comprehensive Assessment System (MCAS, 2011), Utah Test Item Pool Service (UTIPS, 2011), and Virginia Standards of Learning (Virginia-SOL, 2010). Questions were selected on the basis of three criteria: 1) alignment with third- and fourth-grade objectives in the state where the study was conducted; 2) representation of a range of question-type difficulties to differentiate students' scores based on fraction knowledge; and 3) incorporation of a variety of representation types including symbolic items (e.g., numerals and operations only), pictorial items (e.g., pictorial models with a written question stem), and combined items (e.g., numerals and operations combined with pictorial models with a written question stem). One form of 27 multiple-choice questions and 3 open-ended questions was compiled for fourth grade and checked for content validity by five experienced elementary school teachers with graduate degrees. In the fall of the academic year prior to the study, these items were piloted with 275 fifth-grade students from 10 elementary schools in six school districts in order to assess the item difficulties and reliability of the measure.

Item difficulties, fit statistics, reliabilities, and separation indices were estimated using BILOG-MG (Zimowski, Muraki, Mislevy, & Bock, 1996). Biserial correlations for these items were good, ranging between .334 and .776 with a mean of .579 (SD = .115), indicating that the items measured the same construct (i.e., knowledge of fourth-grade fraction concepts). The measure had a high reliability of .8837, showing that it measured knowledge of fourth-grade

fractions consistently across the sampled population. A good distribution of fraction items of different difficulty levels—necessary to differentiate knowledgeable students from less knowledgeable students—was demonstrated by the range of the item difficulties, from about one standard deviation at both ends (–1.198 to .913) with a mean of –.114 (SD = .676). Principles of item response theory were used to construct two forms roughly similar to one another in content and difficulty, but with different sets of items; this was done in order to prevent test-retest effects. All items were then placed on one form and piloted with students, thus linking the original forms to each other using common person equating. Item difficulties on all forms (pre-test, post-test, and delayed post-test) were subsequently calibrated with this entire sample. Each form contained seventeen multiple-choice items with five common linking items and two open-response items. Third-grade tests were created in the same way and piloted with groups of fourth-grade students who were not participating in the study. Procedures for developing the third-grade tests produced similar results.

Observation ethograms. Researchers documented instruction and the use of representations in all the classrooms. Three observers collectively observed 70% of the lessons using a modified ethogram protocol. Ethograms are instruments traditionally used by animal behavior researchers to accurately and efficiently describe the frequency and duration of behaviors made by a species observed in the field, without any subjective evaluation of these observed behaviors (e.g., MacNulty et al., 2007). This instrument provides a cohesive inventory of behavioral patterns describing what a particular species spends its time doing in a studied environment. Here, we used an ethogram adapted for naturalistic classroom observations of humans. At 5-minute intervals throughout the observation of a lesson, observers recorded the types of representations used by teachers and students. For example, observers recorded information on the presentation of mathematical content, terminology, mathematical procedures; use of pictorial, symbolic, physical, and virtual manipulative models; and students' access to physical manipulatives (i.e., passive group viewing or active individual manipulation). VM teachers documented their use of different types of models and students' access to virtual manipulatives via instructor logs, which were subsequently coded and converted to an ethogram protocol as well. These ethograms, in sum, provided a quantitative measure of the students' exposure and access to various fraction concepts, terminology, and types of representation in each of the PM and VM treatment classrooms. These tabulated data provide the basis for the subsequent analyses described below.

RESULTS

The results that follow are organized around our main question and sub-questions. We first present an analysis of the overall mathematics achievement results by grade level to answer the overarching question which focused on possible differences in overall achievement, learning, and retention in third- and fourth-grade

VM and PM treatment classrooms. Then, we address the sub-questions, focusing on the possible mediating variables of *objective ability*, mathematics *content test question type*, time spent in different *instructional configurations*, time spent using each type of fraction *representation*, and interaction effects between instructional configurations by fraction representation.

Learning and Retention

Our overarching research question was: Are there differences in overall achievement on fraction learning and retention between third- and fourth-grade classrooms using virtual fraction applets in a computer lab (VM) and those using texts and physical manipulatives in a regular classroom (PM), as indicated by scores on the pre-test, post-test, and delayed post-tests? We calculated "learning" scores for each student by subtracting the pre-test score from the post-test 1 score. The greater the learning score, the more learning of fraction material between the pre-test and post-test 1. We next calculated a "retention" score to describe the amount of fraction material retained between post-test 1 and post-test 2. The retention variable was calculated by subtracting each student's post-test 2 score from the post-test 1 score. A negative retention score indicates that information was lost between post-tests. Pre-test, post-test 1, and post-test 2 scores for each treatment group in all third-grade classrooms are presented in Table 1 and in all fourth-grade classrooms in Table 2.

Repeated measures analysis of variance (ANOVA) were used to test the null hypothesis for third grade that any obtained differences in test scores both across time (e.g., pre-test, post-test 1, & post-test 2) and across experimental group (e.g., VM & PM) were due to chance. We identified a significant effect of test (pre-test: Mean = 60.37, SD = 20.7; Post-Test 1: Mean = 70.57, SD = 19.62; Post-Test 2: Mean = 62.12, SD = 20.36) ($F(1.975, 302.133) = 8.896$, $MSE = 4052.169$, $p < .000$, $partial\ \eta^2 = .055$), driven by a significant increase in scores from the pre-test to post-test 1 (Bonferroni $p < .001$), followed by a drop in scores from post-test 1 to post-test 2 that approached significance (Bonferroni $p = .054$). There was no difference in scores between pre-test and post-test 2 (Bonferroni $p > .05$). There was no significant difference between the average test scores in VM and PM groups ($F(1, 160) = .694$, $MSE = 588.51$, $p = .41$, $partial\ \eta^2 = .004$). Average pre-test- post-test 1, and post-test 2 scores for each treatment group in third-grade classrooms are shown in Figure 5. The main effect of time is evident in the significant rise in third-grade test scores between pre-test and post-test 1 (p < .001), and nearly significant drop in test scores between post-test 1 and post-test 2 (p = .054). While slight differences in test score averages existed between PM and VM groups within each test, these differences were not statistically significant.

An additional repeated measures ANOVA was conducted for fourth grade. Similar to the third-grade group, results identified a significant main effect of test (pre-test: Mean = 47.97, SD = 19.03; post-test 1: Mean = 65.65, SD = 21.65; post-test 2: Mean = 48.24, SD = 18.58), $F(2, 364) = 100.02$, $MSE =$

| Table 1. Mean (SD) Third-Grade Students Performance: Teacher × Treatment Group × Test | | | | | | |
|---|---|---|---|---|---|
| **Teacher** | **Treatment Group** | **Pre-Test %** | **Post-Test 1 %** | **Post-Test 2 %** | **Learning Avg.** | **Retention Avg.** |
| Mrs. Alpha | VM | 31.81 (12.53) | 72.72 (19.07) | 43.94 (14.56) | 40.91 (24.89) | −28.79 (23.29) |
| | PM | 50.90 (20.93) | 85.45 (8.13) | 50.90 (26.19) | 34.55 (19.71) | −34.55 (24.39) |
| Mrs. Bravo | VM | 69.70 (13.63) | 85.86 (8.01) | 72.73 (7.87) | 16.16 (13.46) | −13.13 (10.27) |
| | PM | 71.71 (21.53) | 79.80 (19.69) | 61.62 (20.21) | 8.10 (11.53) | −18.20 (12.02) |
| Mr. Charlie | VM | 60.61 (18.32) | 75.00 (11.05) | 54.55 (15.98) | 14.40 (14.22) | −20.45 (16.95) |
| | PM | 48.95 (22.12) | 74.13 (19.58) | 57.34 (21.44) | 25.17 (22.31) | −16.78 (16.94) |
| Mrs. Delta* | VM | 56.06 (19.53) | 59.10 (20.26) | 58.57 (25.37) | 3.03 (25.52) | −.51 (23.01) |
| | PM | 58.90 (21.21) | 67.99 (17.52) | 56.52 (19.56) | 9.10 (18.99) | −11.46 (21.27) |
| Mrs. Echo | VM | 67.27 (19.73) | 72.72 (18.68) | 69.10 (18.77) | 5.45 (15.56) | −3.64 (13.68) |
| | PM | 67.27 (24.26) | 75.45 (25.37) | 71.81 (18.40) | 14.54 (12.27) | −3.63 (23.93) |
| Mrs. Foxtrot | VM | 63.64 (17.52) | 69.32 (18.78) | 72.72 (13.74) | 5.68 (10.79) | 3.41 (8.32) |
| | PM | 66.94 (16.40) | 78.51 (13.65) | 72.72 (11.49) | 11.57 (18.67) | −5.78 (11.69) |
| Mrs. Golf | VM | 74.54 (13.41) | 70.00 (10.54) | 76.40 (13.68) | −4.54 (10.71) | 6.36 (14.87) |
| | PM | 58.33 (21.39) | 46.97 (22.21) | 54.54 (24.51) | −11.36 (15.07) | 7.57 (19.31) |
| Total | VM | 60.51 (16.38) | 72.10 (15.20) | 63.99 (15.71) | 11.58 (16.45) | −8.10 (15.77) |
| | PM | 59.52 (21.12) | 72.61 (18.02) | 60.78 (20.26) | 13.09 (16.93) | −11.83 (18.51) |

Note: The asterisks in Tables 1 and 2 indicate teachers who taught more than one class.

Table 2. Fourth-Grade Students Performance: Teacher × Treatment Group × Test						
Teacher	**Treatment Group**	**Pre-Test %**	**Post-Test 1 %**	**Post-Test 2 %**	**Learning Avg.**	**Retention Avg.**
Mr. Hotel*	VM	50.98 (18.63)	61.06 (22.51)	42.85 (19.25)	10.08 (16.12)	−18.21 (17.34)
	PM	52.45 (17.93)	60.78 (19.13)	40.68 (16.63)	8.33 (17.25)	−20.09 (16.89)
Mrs. India*	VM	49.85 (20.00)	67.50 (20.93)	54.34 (14.98)	17.64 (20.46)	−13.16 (20.45)
	PM	49.41 (21.62)	69.11 (19.91)	46.47 (16.84)	19.40 (22.92)	−22.64 (16.88)
Mrs. Juliet	VM	47.05 (20.88)	67.37 (18.51)	45.45 (15.07)	20.32 (20.60)	−21.92 (13.94)
	PM	47.89 (17.39)	60.51 (23.28)	38.65 (16.65)	12.60 (15.27)	−21.84 (11.86)
Mrs. Kilo	VM	47.05 (20.65)	57.98 (19.92)	45.37 (22.19)	10.92 (12.44)	−12.60 (20.77)
	PM	43.53 (24.04)	58.82 (25.41)	40.00 (17.71)	15.29 (15.23)	−18.82 (21.79)
Mrs. Lima	VM	35.94 (16.49)	67.32 (30.86)	57.51 (22.74)	31.37 (24.07)	−9.80 (12.82)
	PM	39.57 (14.43)	69.51 (19.65)	55.08 (14.94)	29.94 (14.03)	−14.43 (11.87)
Mrs. Mike	VM	61.17 (18.84)	80.58 (25.27)	59.41 (24.08)	19.41 (17.54)	−21.17 (9.86)
	PM	50.98 (19.32)	75.49 (20.96)	64.21 (15.35)	24.50 (20.19)	−11.27 (13.35)
Mrs. November	VM	42.64 (16.47)	59.80 (20.80)	45.09 (20.42)	17.15 (23.32)	−14.70 (14.94)
	PM	41.17 (14.18)	67.64 (19.18)	51.96 (13.23)	53.47 (21.35)	−15.68 (17.25)
Total	VM	47.81 (18.58)	65.94 (23.37)	50.01 (20.04)	18.13 (19.22)	−15.94 (16.23)
	PM	46.43 (18.42)	65.98 (20.39)	48.15 (15.68)	19.55 (18.03)	−17.83 (15.20)

Note: The asterisks in Tables 1 and 2 indicate teachers who taught more than one class.

Figure 5. Third-Grade Performance: Treatment Group × Test
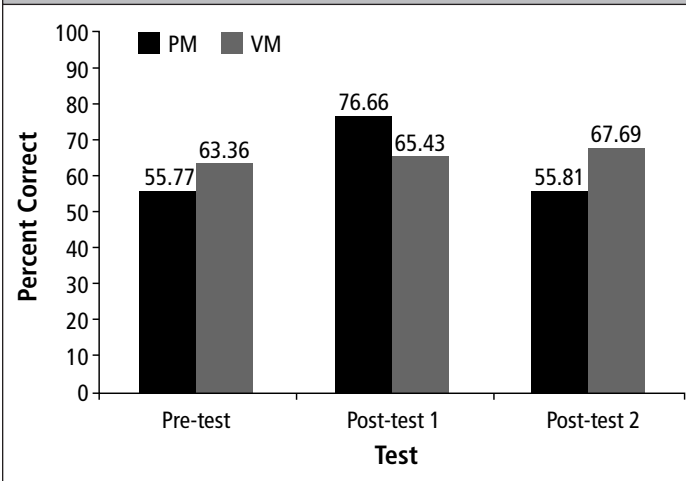
16295.68, $p < .000$, *partial* $\eta^2 = .36$, indicating that the average scores differed significantly across tests. This effect was driven by a significant difference in post-test 1 scores compared to pre-test and post-test 2 (Bonferroni $p < .001$ for both comparisons). There was no difference between pre-test and post-test 2 scores (Bonferroni $p > .05$). There was no significant difference between the average test scores in VM and PM groups ($F(1, 195) = .002$, $MSE = 1.72$, $p = .95$, *partial* $\eta^2 = .00$). Average pre-test-post-test 1, and post-test 2 scores for each treatment group in fourth-grade classrooms are shown in Figure 6. The main effect of time is again evident in the significant rise in test scores between pre-test and post-test 1 ($p < .001$), and subsequent significant drop in test scores between post-test 1 and post-test 2 ($p < .001$). Again, for fourth grade the differences in test scores between PM and VM groups were not statistically significant.

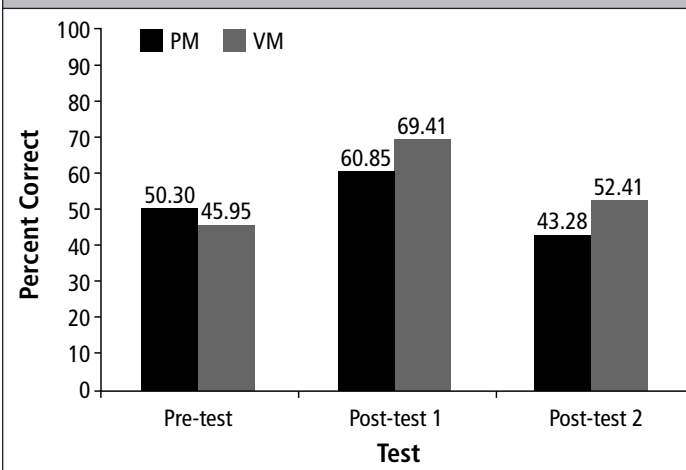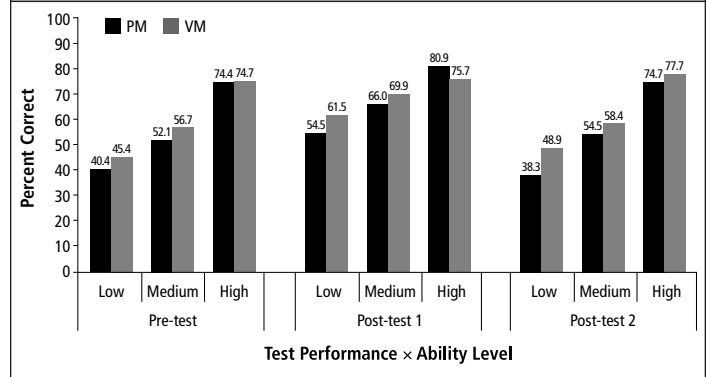**Figure 6.** Fourth-Grade Performance: Treatment Group × Test

Figure 7. Third-Grade Performance: Group × Objective Ability × Test

Effects of Objective Ability

Our first subquestion was: Does student *objective ability,* based on pre-test student score in relation to class pre-test average, predict fraction achievement, learning, and retention in VM or PM classrooms? In third grade, we identified a significant interaction between test and objective ability ($F(1.975, 302.133) = 6.534$, $MSE = 2976.379$, $p = .002$, partial $\eta^2 = .041$), indicating that the scores for students objectively rated as "High" (Mean = 76.39, SD = 15.49) increased more dramatically from pre-test to post-test 1 than students rated "Low" (Mean = 59.64, SD = 17.88) and "Medium" (Mean = 48.21, SD = 20.02). Additionally, objectively rated "High" students maintained higher performance from post-test 1 to post-test 2 than "Low" or "Medium" students (see Figure 7). No other comparisons were significant.

A similar analysis in fourth grade identified a significant main effect of objective ability (High: Mean = 60.75, SD = 2.09; Medium: Mean = 52.97, SD = 1.66; Low: Mean = 49.88, SD = 2.73), $F(2, 182) = 6.275$, $MSE = 4724.45$, $p = .002$, partial $\eta^2 = .065$, with significant individual comparisons between objectively rated High and Low students (Bonferroni $p = .006$), and between objectively rated High and Medium students (Bonferroni $p = .012$), but not between Medium and Low students (Bonferroni $p > .05$). No other main effects were identified (Figure 8).

Figure 8. Fourth-Grade Performance: Group × Objective Ability × Test
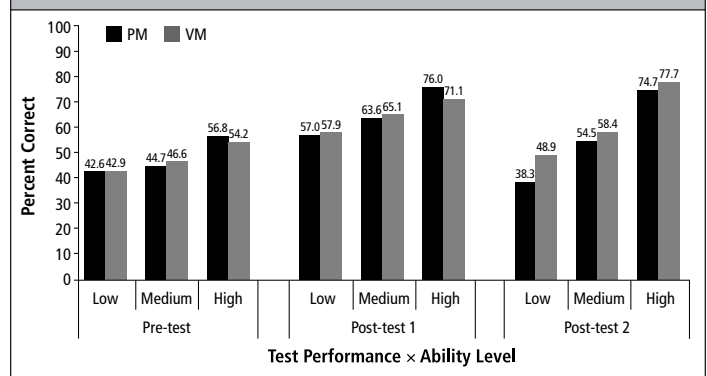
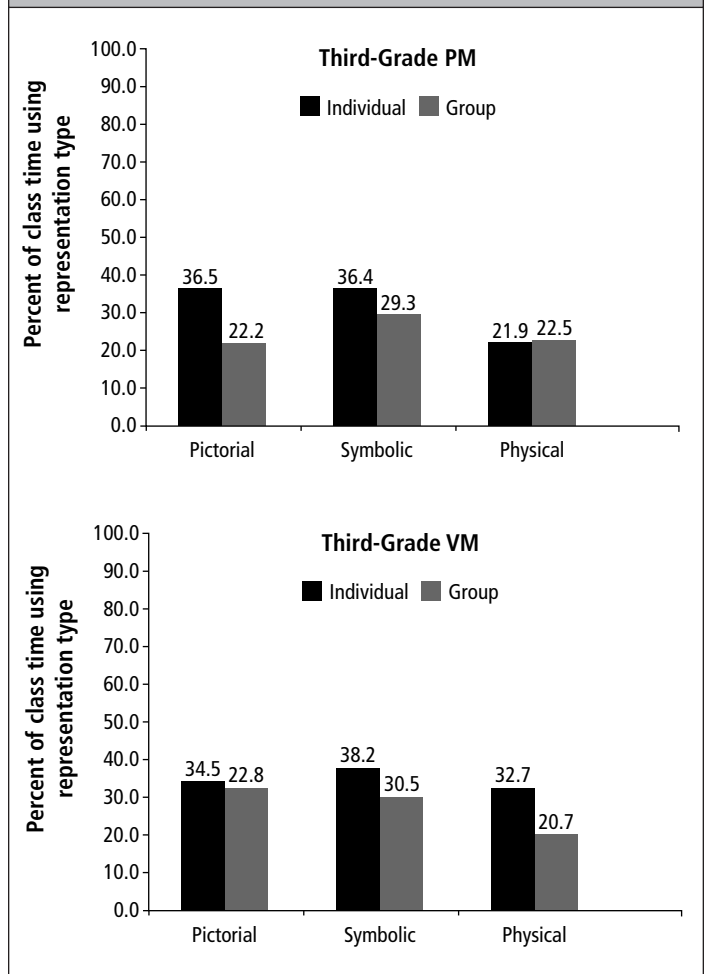## Effects by Content Question Type: Visual, Symbolic, and Combined

Our next sub-question was: Is any impact of virtual manipulative use on students' fraction achievement modulated by mathematics *test question type* (e.g., pictorial, symbolic, and combined)? No third-grade classroom provided purely symbolic question types, so only pictorial and combined questions types were analyzed for this sample. In third grade, post-test 1, we identified a significant main effect of question type ($F(1, 154) = 33.56$, *MSE* = 10650.06, *H-F p* < .001, *partial* $\eta^2 = .179$), driven by greater overall performance on combined ($\mu = 78.05$) compared to pictorial ($\mu = 66.29$) question types. This difference was not observed for post-test 2 (*H-F p* > .05). No other analyses showed statistically significant differences. This pattern of results indicates that combined question types revealed third-grade students' short-term learning (i.e., post-test 1) of fraction material to a greater degree than pictorial question types, but no difference was found between the two question types in terms of long-term retention.

Pictorial, symbolic, and combined questions types were analyzed for fourth grade. A similar trend as that identified in third grade emerged for fourth grade: Repeated measures ANOVA identified a significant main effect of question type ($F(2, 328.08) = 16.38$, *MSE* = 9125.42, *H-F p* < .001, *partial* $\eta^2 = .079$). Pairwise t-test comparisons with Bonferroni adjustment identified significant differences between combined ($\overline{\chi} = 70.06$) and pictorial ($\overline{\chi} = 57.55$) question types (*p* < .001), as well as between combined and symbolic ($\overline{\chi} = 61.97$) question types (*p* < .001), but not between pictorial and symbolic question types (*p* > .05). However, unlike our third-grade group, the difference in performance between question types persisted throughout the delayed post-test ($F(2, 352.154) = 62.522$, *MSE* = 25313.102, *H-F p* < .001, *partial* $\eta^2 = .246$). Follow-up pairwise comparisons with Bonferroni adjustment identified significant differences between pictorial ($\overline{\chi} = 34.52$) and combined ($\overline{\chi} = 54.40$) question types (*p* < .001), as well as between pictorial and symbolic ($\overline{\chi} = 52.38$) question types (*p* < .001), but not between combined and symbolic question types (*p* > .05). Thus, in fourth grade, the effect of question types began at post-test 1 and, unlike the third-grade sample this effect persisted into long-term retention on post-test 2.

Effects by Instructional Configuration: Group vs. Individual

The next sub-question was: Does percentage of class time spent in different *instructional configurations* (e.g., *groups* vs. *individually*) differ between VM and PM groups? We conducted a repeated measures t-test to compare time students spent engaged in whole group versus individual instruction for each representation (pictorial, symbolic, and physical/virtual manipulative). The third-grade repeated measures analyses found no significant difference for the PM group between percentage of time students used representations in whole group instruction compared to when they worked individually (Figure 9). However, a trending difference was found for the VM group [$t(6) = 2.221$, *p* = .068], with students spending
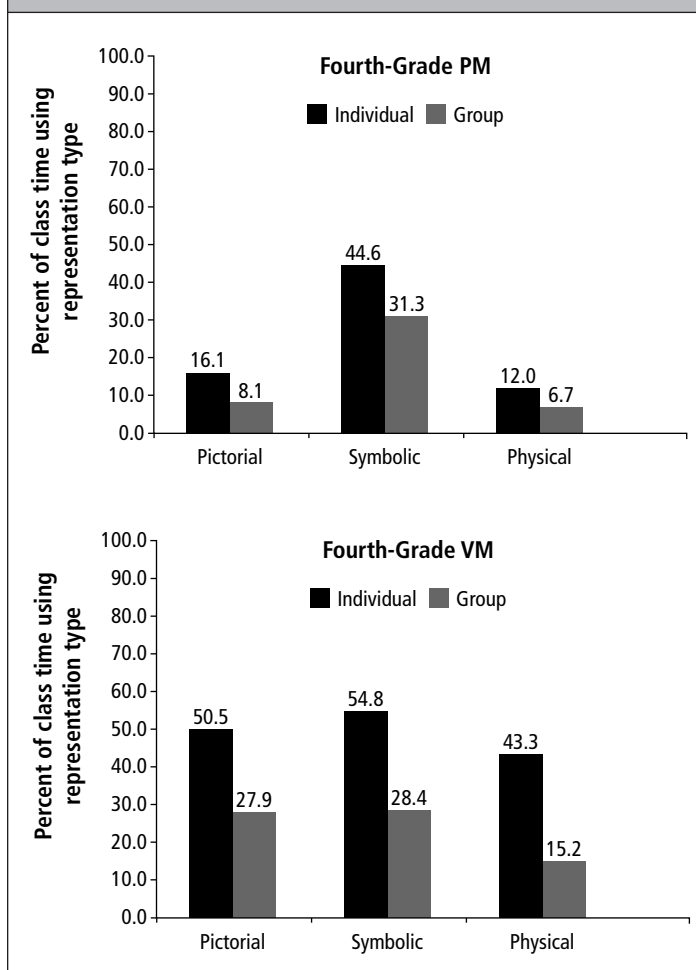
more class time working individually (M = 38.61%) compared to participating in whole group instruction (M = 18.14%). This trend is not surprising because the students using virtual manipulatives in the computer lab frequently worked individually on tasks at their own computers.

In fourth grade, the repeated-measures analyses found no significant difference for the PM group between percentage of time students used representations working in whole groups compared to percentage of time they worked individually (Figure 10). Conversely, a significant difference was found for the VM group [$t(6) = 3.074$, *p* = .022], which spent significantly more class time working individually (M = 38.9%) compared to working in a whole group (M = 16.99%). Again, this significant difference was expected due to the individual nature of using virtual manipulative tools in a computer lab classroom. No further statistical differences were found.

Effects by Fraction Representation Type: Pictorial, Symbolic, Manipulative

The next sub-question was: Does percentage of class time spent using each type of fraction *representation* (e.g., pictorial, symbolic, manipulative type) differ within and between VM and PM

Figure 10. Comparing Time Spent Individually or in Group Instruction for Fourth-Grade PM (top) and VM (bottom) Groups

Interaction Effects Between Instructional Configuration and Fraction Representation

The final sub-question asked: Are there interaction effects of class time spent in different instructional configurations by fraction representation? In other words, as these representations were often used both individually by each student as well as in a group setting, we were interested in whether or not the duration of use of each representation differed across individual and group settings. In our third-grade sample, follow-up pairwise comparisons to the repeated measures ANOVA described in the previous section, with Bonferroni adjustments, identified significant individual differences between time spent using pictorial representations individually ($\bar{\chi} = 35.45$) and VM-PM manipulative types in a group setting ($\bar{\chi} = 21.59$) ($p = .022$), as well as between VM-PM manipulative types and symbolic representations used in a group setting ($\bar{\chi} = 29.89$) ($p = .005$). No other comparisons were significant for third grade.

Similar analyses conducted in fourth grade identified multiple significant differences between representation usage. Symbolic representations used individually were far-and-away the most often used representation type ($\bar{\chi} = 50.91$) in fourth grade, therefore many of the significant pairwise comparisons emerged as a result of direct comparison with this representation type. For instance, time spent using VM-PM manipulative types individually ($\bar{\chi} = 28.4$), pictorial representations in a group setting ($\bar{\chi} = 18.52$), and VM-PM manipulative types in a group setting ($\bar{\chi} = 10.93$) all differed significantly from time spent using symbolic representations individually ($p < .01$). Similarly, as VM-PM manipulative types in a group setting were the least overall used representation ($\bar{\chi} = 10.93$) in fourth grade, every comparison other than pictorial representations used in a group setting differed significantly from it ($p < .05$). These results indicate that the amount of time dedicated to each representation often differs depending on instructional configuration (e.g., individual or group). In other words, not only does representation usage differ across the teaching method being used (e.g., VM or PM), but this pattern of usage also depends on the instructional configuration (e.g., individual or group).

LIMITATIONS

Even in studies in which students are randomly assigned and treatments are clearly described, we acknowledge that these are not laboratory settings. Students' responses and actions may unintentionally be influenced by a context in which they are not participating in mathematics learning with their regular teacher, their own beliefs about participating in research, and teacher and researcher expectations. "We have also long known, both from experiments and everyday experience, how subjects' behaviors are affected by expectation, context, and measurement procedures. The notion that there can be 'neutral' methods for gathering data has been refuted decisively" (Ericsson & Simon, 1981, p. 17).

Another limitation was the possibility of teacher effects which we attempted to reduce by ensuring that the instructors in the 17

classrooms? Repeated measures ANOVA were conducted on the observation data collected within each classroom, which indicates the amount of time each representation was used in both settings. For third-grade classrooms, a significant main effect of representation type was identified ($F(5, 38.98) = 3.11$, $MSE = 703.987$, $H\text{-}F\ p = .034$, partial $\eta^2 = .206$), which indicates that the amount of time spent using each representation differed significantly. No other comparisons were significant in third grade.

Identical comparisons were then conducted on the fourth-grade sample: Here, again, repeated measures ANOVA identified a significant main effect of representation type ($F(5, 48.302) = 15.08$, $MSE = 4368.976$, $H\text{-}F\ p < .001$, partial $\eta^2 = .519$). Also, a significant interaction between representation type and classroom type (VM or PM) was identified ($F(5, 48.302) = 3.426$, $MSE = 992.281$, $H\text{-}F\ p = .020$, partial $\eta^2 = .197$). Similarly, a significant main effect of classroom type (VM or PM) was identified ($F(1, 14) = 17.96$, $MSE = 6198.52$, $p = .001$, partial $\eta^2 = .562$), driven by greater overall use of each type of representation in VM compared to PM classrooms in fourth grade. These results indicate that the amount of time dedicated to each representation differed depending on PM or VM classroom representations.

VM groups and the 17 PM groups (a total of 34 instructional groups) had similar backgrounds and teaching experience, taught the same mathematics content daily, and adhered to a set of guidelines for the use of physical and virtual manipulatives. In this study with a large number of classrooms and different instructors, we attempted to address the potential impact of different teachers on student performance by ensuring that lesson content was the same on each day of the mathematics lessons and by observing over 70% of the classroom instruction. Furthermore, as using either physical or virtual manipulatives produced similar student learning of fraction concepts, results suggest that even when there could potentially be differences between a large number of instructors, student achievement remains equal across these different mathematics instructional modalities. Repeated measures ANOVA indicated that test scores (i.e., pre-test, post-test 1, post-test 2) did differ significantly across individual classrooms involved in the study ($F(13,54) = 7.79$, $MSE = 389.0$, p .001). However, despite this difference across classrooms, an insignificant classroom × instruction group PM vs. VM) interaction indicated that students' instructional group did not differentially mediate test score outcomes across classrooms that differed in test performance ($F(13,54) = 1.77$, $MSE = 88.7$, $p > .05$).

An additional analysis of the potential demographic predictors behind these achievement results was beyond the scope of this paper. A separate paper, using a variety of demographic variables (e.g., socio-economic status, English language learner status, and gender) in a linear regression analysis, examines these potential predictors (Moyer-Packenham, Jordan, et al., 2013). The results of the additional analysis revealed that fewer demographic predictors of student performance (e.g., socio-economic status, English language learner status, and gender) existed during fraction instruction using virtual manipulatives. When instructors used virtual manipulatives, there was an equalizing effect on achievement in third and fourth grade, in that fewer demographic factors were influential in the VM groups.

DISCUSSION

This study utilized the following design elements: 1) large numbers of student participants; 2) within-class random-assignment of students to treatment groups; 3) delayed post-tests to measure retention effects; 4) observations to document fidelity of instructional treatments, consistency of mathematical content, and representation use; and, 5) instrument development using IRT. The overarching finding was that, when uniquely combining multiple elements of a rigorous research design, no differences in overall achievement or fraction learning and retention emerge in third- and fourth-grade classrooms where virtual manipulatives are compared with classrooms using physical manipulatives and text-based materials. Essentially, when students were exposed to the same mathematics content and the same types of representations and spent about the same amount of time learning mathematics content in individual or group configurations, student performance on mathematics assessment measures of learning and retention were equal.

Within the context of this broad research finding, important results also emerged from our sub-questions. First, student objective ability, based on pre-test student score in relation to class pre-test average, predicted fraction achievement, learning, and retention in both VM and PM classrooms. Specifically, in both third and fourth grades, students with high objective ability showed a greater increase from pre-test to post-test 1 than did students with low and medium objective abilities. In addition, scores for students with high objective ability remained higher between post-test 1 and the delayed post-test than did scores for students with low and medium objective abilities. Therefore, student objective ability was a reliable predictor of fraction learning and performance in multiple classroom format modalities. Students' pretest scores, in relation to the class pretest average, are related to their learning and retention on the fraction tests administered in this study; this held for both VM and PM classroom modalities.

Secondly, we found that the impact of virtual manipulative use on students' fraction achievement was modulated by mathematics content test question type. The impact of question type on student achievement was of longer duration for fourth graders than for third graders. For fourth grade, as with third grade, students did better on combined question types, but this effect persisted longer for fourth graders. On post-test 2, fourth-grade students also performed better on symbolic than pictorial question types.

The current study shows that percentage of class time spent using fraction representations in groups vs. individually differed between VM and PM classrooms. Specifically, in third grade there was a trend to spend more class time using representations individually in VM but not in PM classrooms, and in fourth grade this difference reached significance. This was not surprising because individual students were at individual computers using virtual manipulatives in a computer classroom. Percentage of class time spent using each type of fraction representation was compared between VM and PM classrooms. In third grade there was no significant difference between percentage of class time spent using each type of representation in VM or PM classrooms. In fourth grade the VM classrooms spent significantly more time than PM classrooms using the manipulative representations. However, the results on time spent individually versus group work and the differences in the use of representations did not produce significant differences in students' learning or retention.

In our final interaction analysis, percentage of class time that students spent using each type of representation differed according to instructional configuration (individual vs. group). For example, in third grade, students spent more group time using symbolic representations than manipulatives. In fourth grade, the amount of time spent using symbolic representations individually was greater than any other representation type/instructional configuration, and time spent using manipulatives in a group was less than any other representation type/instructional configuration. Thus, amount of time using each representation differed significantly according to instructional configuration (e.g., individual or group).

IMPLICATIONS FOR PRACTICE

With very few significant differences between the VM and PM groups, what are the implications for classroom practice? First, we hope our results will help to put to rest the persistent question: Which is better—physical or virtual manipulatives? These results, based on a rigorous research design, demonstrate that using either physical or virtual manipulatives produce similar student achievement for third- and fourth-grade students learning fraction concepts. Overall, our results reveal that experienced instructors can use different instructional modalities for mathematics instruction and produce similar achievement results.

Another implication for practice is that classroom teachers can use pretests to identify students who may have particular difficulties during a mathematics unit of study, and provide specific RtI intervention during the unit to support those students identified as "Low."

The importance of students' facility with representations and tools was confirmed. For example, two mathematical practices in the *Common Core State Standards* (e.g., model with mathematics and use appropriate tools strategically) are integral to students' learning. A lack of familiarity with different representations can negatively influence students' understanding. In this study, students learned to use physical or virtual tools, and both instructional media supported their short-term learning from the pre- to the post-tests.

Finally, these results suggest that there are a variety of instructional configurations and representation that can produce learning gains. For example, students in a VM classroom have a greater opportunity to learn from representations individually (as opposed to sharing these representations in groups); similarly, students in VM classrooms may have greater exposure to pictorial representations of fractions, while students in PM classrooms have greater exposure to working with symbolic representations. However, all of the groups experienced quite a loss of learning based on the retention scores on the delayed post-test. With various instructional modalities, it was still difficult to retain knowledge and learning of fraction concepts for these third- and fourth-grade students. An important avenue for future research is to determine how to strengthen and solidify learning so that students not only learn the fraction concepts, but also retain the mathematics concepts.

CONCLUSION

This study addressed the need for large-scale, random-assignment, school-based, delayed effects research on virtual manipulatives compared with instruction using texts and physical manipulatives. Results revealed no significant overall differences in achievement between the VM and PM treatment groups when multiple elements of a quasi-experimental design were utilized. Additional results showed that student objective ability predicted fraction achievement in both VM and PM classrooms; the impact of virtual manipulative use on students' fraction achievement can be modulated by mathematics content test question type (e.g., symbolic vs.

pictorial); percentage of class time spent using fraction representations in groups vs. individually can differ between VM and PM classrooms; and percentage of class time spent using each type of fraction representation (e.g., pictorial vs. symbolic) can differ between VM and PM classrooms, potentially providing differential opportunities to learn through each type of instructional modality.

References

Arcavi, A., & Hadas, N. (2000). Computer mediated learning: An example of an approach. *International Journal for Computers for Mathematics Learning, 5*, 25–45.

Ball, S. (1988). Computers, concrete materials and teaching fractions. *School Science and Mathematics, 88*, 470–475.

Ball, D. L. (1992). Magical hopes: Manipulatives and the reform of math education. *American Educator, 16*(2), 14–18, 46–47.

Berlin, D. & White, A. (1986). Computer simulations and the transition from concrete manipulative of objects to abstract thinking in elementary school mathematics. *School Science and Mathematics, 86,* 468–479.

Bruner, J. S. (1960). *The process of education*. Cambridge, MA: Harvard University Press.

Bruner, J. S. (1986). *Actual minds, possible worlds*. Cambridge MA: Harvard University Press.

Burns, B. A., & Hamm, E. M. (2011). A comparison of concrete and virtual manipulative use in third and fourth grade mathematics. *School Science and Mathematics, 111*(6), 256–261.

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

Clements, D. H., & Battista, M. T. (1989). Learning of geometric concepts in a logo environment. *Journal for Research in Mathematics Education, 20*(5), 450–467.

Clements, D. H., Battista, M. T., & Sarama, J. (2001). Logo and geometry [Monograph]. *Journal for Research in Mathematics Education,10,* 1–177.

Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the *Building Blocks* project. *Journal for Research in Mathematics Education, 38*(2), 136–163.

Cobb, P. (1995). Cultural tools and mathematical learning: A case study. *Journal for Research in Mathematics Education, 26*(4), 362–385.

Dienes, Z. P. (1969). *Building up mathematics*. London, UK: Hutchison Education.

Dinardi-Besterman, D. (1992). *A comparison of two computer-assisted methods for teaching first grade students multidigit addition and subtraction* (Master's thesis). San Jose State University, San Jose, CA. Retrieved from http://scholarworks.sjsu.edu/etd_theses/452.

Duffy, T. M., & Cunningham, D. J. (1996). Constructivism: Implications for the design and delivery of instruction. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 170–198). New York, NY: Simon and Schuster Macmillan.

Ericsson, K., & Simon, H. (1981). Sources of evidence on cognition: A historical overview. In T. Merluzzi, C. Glass, & M. Genest (Eds.), *Cognitive assessment* (pp. 16–51). New York, NY: Guilford.

Fuson, K. C., & Briars, D. J. (1990). Using a base-ten blocks learning/teaching approach for first- and second-grade place-value and multidigit addition and subtraction. *Journal for Research in Mathematics Education, 21*(3), 180–206.

Goldin, G. A. (2003). Representation in school mathematics: A unifying research perspective. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 275–285). Reston, VA: NCTM.

Hauptman, H. (2010). Enhancement of spatial thinking with Virtual Spaces 1.0. *Computers and Education, 54*(1), 123–135.

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K. C., Wearne, D., Murray, H., Olivier, A., & Human, P. (1997). *Making sense: Teaching and learning mathematics with understanding*. Portsmouth, NH: Heinemann.

Lesh, R., & Doerr, H. M. (2003). Foundations of a models and modeling perspective on mathematics teaching, learning and problem solving. In R. Lesh & H. M. Doerr (Eds.), *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning and teaching* (pp. 3–33). Mahwah, NJ: Erlbaum.

Lin, C. (2010). Web-based instruction on pre-service teachers' knowledge of fraction operations. *School Science and Mathematics, 110*(2), 59–71.

Martin, T., & Lukong, A. (2005, April). *Virtual manipulatives: How effective are they and why?* Paper presented at the American Educational Research Association Annual Meeting, Montreal, Canada.

Massachusetts Comprehensive Assessment System. (2011). Massachusetts Comprehensive Assessment System website. Retrieved from http://www.doe.mass.edu/mcas/.

MacNulty, D.R., Mech, D.L., & Smith, D.W. (2007). A proposed ethogram of large-carnivore predatory behavior, exemplified by the wolf. *Journal of Mammalogy, 88*, 595–605.

Meira, L. (1998). Making sense of instructional devices: The emergence of transparency in mathematical activity. *Journal for Research in Mathematics Education, 29*(2), 121–142.

Melideo, S., & Dodson, F. (2009). Directly from pedagogy to practice: Incorporating virtual manipulatives in the elementary mathematics classroom. *The Teacher Educator's Journal, 16*, 28–35.

Mendiburo, M., & Hasselbring, T. (2011, March). *Technology's impact on fraction learning: An experimental comparison of virtual and physical manipulatives*. Paper presented at the Society for Research on Educational Effectiveness, Washington, DC.

Moreno, R., & Mayer, R. E. (1999). Multimedia-supported metaphors for meaning making in mathematics. *Cognition and Instruction, 17*(3), 215–248.

Moyer, P. S. (2001). Are we having fun yet? How teachers use manipulatives to teach mathematics. *Educational Studies in Mathematics, 47*(2), 175–197.

Moyer, P. S., Bolyard, J. J., & Spikell, M. A. (2002). What are virtual manipulatives? *Teaching Children Mathematics, 8*(6), 372–377.

Moyer, P. S., & Jones, M. G. (2004). Controlling choice: Teachers, students, and manipulatives in mathematics classrooms. *School Science and Mathematics, 104*(1), 16–31.

Moyer-Packenham, P. S., Jordan, K, Baker, J., Westenskow, A., Rodzon, K., Anderson, K., & Shumway, J. (2013, April). *Hidden predictors of achievement: The equalizing effect of virtual manipulatives for mathematics instruction*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, California.

Moyer-Packenham, P. S., Salkind, G. W., Bolyard, J., & Suh, J. M. (2013). Effective choices and practices: Knowledgeable and experienced teachers' uses of manipulatives to teach mathematics. *Online Journal of Education Research, 2*(2), 18–33.

Moyer-Packenham, P. S., & Suh, J. M. (2012). Learning mathematics with technology: The influence of virtual manipulatives on different achievement groups. *Journal of Computers in Mathematics and Science Teaching, 31*(1), 39–59.

Moyer-Packenham, P. S., & Westenskow, A. (2012, April). *Effects of virtual manipulatives on student achievement and mathematics learning*. Paper presented at the Annual Meeting of the American Educational Research Association, Vancouver, British Columbia, Canada.

National Assessment of Educational Progress. (2011). National Center for Education Statistics website. Retrieved from http://nces.ed.gov/nationsreportcard/about/naeptools.aspx.

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common Core State Standards*. Washington, DC: Author.

Nute, N. E. (1997). *The impact of engagement activity and manipulatives presentation on intermediate mathematics achievement, time-on-task, learning efficiency, and attitude*. (Unpublished doctoral dissertation). The University of Memphis, Memphis, Tennessee. Available from ProQuest Dissertations and Theses database. (UMI No. 9807175)

Parham, J. L. (1983). A meta-analysis of the use of manipulative materials and student achievement in elementary school mathematics. *Dissertation Abstracts International, 44A*, 96.

Piaget, J. (1952). *The child's conception of number*. New York, NY: Humanities Press.

Pleet, L. J. (1991). *The effects of computer graphics and Mira on acquisition of transformation geometry concepts and development of mental rotation skills in grade eight*. (Unpublished doctoral dissertation). Oregon State University, Corvallis, Oregon.

Prigge, G. R. (1978). The differential effects of the use of manipulative aids on the learning of geometric concepts by elementary school children. *Journal for Research in Mathematics Education, 9*(5), 361–367.

Raphael, D., & Wahlstrom, M. (1989). The influence of instructional aids on mathematics achievement. *Journal for Research in Mathematics Education, 20*(2), 173–190.

Reimer, K., & Moyer, P. S. (2005). Third graders learn about fractions using virtual manipulatives: A classroom study. *Journal of Computers in Mathematics and Science Teaching, 24*(1), 5–25.

Schoenfeld, A. H. (1983). Beyond the purely cognitive: Belief systems, social cognitions, and metacognitions as driving forces in intellectual performance. *Cognitive Science, 7*, 329–363.

Schoenfeld, A., Smith, J., & Arcavi, A. (1993). Learning: The microgenetic analysis of one student's evolving understanding of a complex subject matter domain. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. IV, pp. 55–175). Hillsdale, NJ: Erlbaum.

Smith, L. (2006). *The impact of virtual and concrete manipulatives on algebraic understanding*. (Unpublished doctoral dissertation). George Mason University, Fairfax, Virginia.

Sowell, E. J. (1989). Effects of manipulative materials in mathematics instruction. *Journal for Research in Mathematics Education, 20*(5), 498–505.

Steen, K., Brooks, D., & Lyon, T. (2006). The impact of virtual manipulatives on first grade geometry instruction and learning. *Journal of Computers in Mathematics and Science Teaching, 25*(4), 373–391.

Suh, J. M., & Moyer-Packenham, P. S. (2007). The application of dual coding theory in multi-representational virtual mathematics environments. In J. H. Woo, H. C. Lew, K. S. Park, & D. Y. Seo (Eds.), *Proceedings of the 31st conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 209–216). Seoul, KR: PME.

Suydam, M. N. (1985). *Research on instructional materials for mathematics*. Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Suydam, M. N., & Higgins, J. L. (1977). *Activity-based learning in elementary school mathematics: Recommendations from research*. Columbus, OH: ERIC Center for Science, Mathematics & Environmental Education, College of Education, Ohio State University.

Terry, M. K. (1995). *An investigation of differences in cognition when utilizing math manipulatives and math manipulative software*. (Unpublished doctoral dissertation). University of Missouri, St. Louis, Missouri Available from ProQuest Dissertations and Theses database. (UMI No. 9536433).

Thompson, P. W. (1985). A Piagetian approach to transformation geometry via microworlds. *Mathematics Teacher*, *78*, 465–471.

Thompson, P. W. (1992). Notations, conventions, and constraints: Contributions to effective uses of concrete materials in elementary mathematics. *Journal for Research in Mathematics Education, 23*(2), 123–147.

Uribe-Florez, L. J., & Wilkins, J. L. M. (2010). Elementary school teachers' manipulative use. *School Science and Mathematics, 110*(7), 363–371.

Utah Test Item Pool Service. (2011). Utah Test Item Pool Service website. Retrieved from https://www.utips.org/#.

Virginia Standards of Learning. (2010). Virginia Department of Education website. Retrieved from http://www.doe.virginia.gov/testing/sol/standards_docs/.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Zbiek, R. M., Heid, M. K., Blume, G. W., & Dick, T. P. (2007). Research on technology in mathematics education: The perspective of constructs. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 2, pp. 1169–1207). Charlotte, NC: Information Age Publishing Inc.

Zimowski, M.F., Muraki, E., Mislevy, R.J., & Bock, R.D. (1996). *BILOG-MG: Multiple-group IRT analysis and test maintenance for binary items*. Chicago, IL: Scientific Software International.

Patricia Moyer-Packenham is a professor and the Director of Mathematics Education and Leadership in the School of Teacher Education and Leadership at Utah State University. Dr. Moyer-Packenham can be reached at patricia.moyer-packenham@usu.edu.

Joseph Baker is a post-doctoral researcher in the Department of Psychology at Utah State University. Dr. Baker can be reached at baker.josephm@gmail.com.

Arla Westenskow is a post-doctoral researcher in the School of Teacher Education and Leadership at Utah State University. Dr. Westenskow can be reached at arlawestenskow@gmail.com.

Katie Anderson is a doctoral candidate in the School of Teacher Education and Leadership at Utah State University. Ms. Anderson can be reached at katie.anderson@aggiemail.usu.edu.

Jessica Shumway is a doctoral student in the School of Teacher Education and Leadership at Utah State University. Ms. Shumway can be reached at jfshumway10@gmail.com.

Kati Rodzon is a former doctoral student in the Department of Psychology at Utah State University. Ms. Rodzon can be reached at krodzon@gmail.com.

Kerry Jordan is an associate professor in the Department of Psychology at Utah State University. Dr. Jordan can be reached at Kerry.jordan@usu.edu.