The Impacts of Intervention Modality on Student Mathematics Operations Knowledge

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Kourtney R. Kromminga

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dr. Robin Codding, Dr. Amanda Sullivan

July 2021

Author Note

The contents of this manuscript were developed under a grant from the U.S. Department of Education, # H325D160016. However, those contents do not necessarily represent the policy of the U.S. Department of Education, and you should not assume endorsement by the Federal Government. Project Officer, Sarah J. Allen, Ph.D.

Correspondence concerning this article should be addressed to Kourtney R. Kromminga, 56 E River Rd., 250 Education Sciences Building, University of Minnesota – Twin Cities, MN 55455. Email: mcnal096@umn.edu

Acknowledgements

Thank you to my committee, including my advisor Dr. Robin Codding who provided me with new research opportunities and always had confidence in my success. To Dr. Amanda Sullivan whose support and guidance helped me stay on track. To Dr. Bodong Chen who has had a genuine interest in these projects, and Dr. Jessie Kember, who has been incredibly supportive throughout my entire career in school psychology. I would also like to acknowledge my undergraduate mentors, Dr. Daniel Houlihan, Dr. Carlos Panahon, and Dr. Shawna Petersen-Brown, your support and confidence got me into this program, and I am forever grateful. To my fellow students, I thank you for the hours of study groups, words of encouragement, research support, and so much more.

I would also like to thank my family and friends without whom, I would not have gotten to this point. To my mom, thank you for showing me how to be strong and independent. To my mother-in-law, thank you for always being there. To my sisters, brother, father-in-law, and brother-in-law, thank you for putting up with long winded stories about school and research projects. Thank you, Judy, for teaching me to advocate for children and for your endless words of wisdom, I hope you are looking down and still proud. Lastly, thank you to my husband, who knew me before graduate school, married me during it, and has stayed with me despite it. Without your daily support, I truly could not have done it. To everyone named above and many more who are not named here, thank you for your unending mentorship, encouragement, love, and confidence.

Dedication

To Judy, I hope that I can be the fierce advocate for children that you once were and learn to approach adversity with the wisdom and courage that you did.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

Many students in the United States struggle with the acquisition of critical mathematics concepts such as number combinations, procedural fluency, computation, and problem solving (Berch & Mazzocco, 2007; Jordan et al., 2010; National Mathematics Advisory Panel [NMAP] 2008). In fact, in 2019, over half of all 4th and 8th grade students did not meet proficiency standards in mathematics on the National Assessment of Educational Performance (NAEP; NCES, 2019). Specifically, 4th grade students are not meeting basic proficiency standards in whole number operations on the numbers and operations subtest of the NAEP (NCES, 2019). In addition to underachievement on national assessments, students in the U.S. consistently perform below students from other developed countries on tests of whole number knowledge (NCES, 2015). Taken together this means that most students in the U.S. are not able to demonstrate competence on tests of whole number operations and this problem has persisted over time (NCES, 2019).

These results are concerning considering that research indicates that mastery of basic number operations (i.e., addition, subtraction, multiplication, and division) is a critical skill for solving more advanced math problems such as those in algebra, rational number computation, and word problem solving (Fuchs et al., 2006; Fuchs et al., 2016, Namkunk et al., 2018; NMAP, 2008). For example, whole number knowledge has been found to be a predictor of fraction achievement and students with low whole number knowledge had more difficulty with basic fraction concepts and procedures (Jordan et al., 2017). Moreover, students' ability to apply algebraic principals is highly dependent on mastery of whole number operations (Ketterlin-Geller & Chard, 2011; Fuchs et al., 2016). Failure to master whole number operations makes it less likely students will attend college and has been linked to less career success and lower incomes (Adelman, 2006; Attewell & Domina, 2008; Dowker, 2005). Additionally, adults with low mathematics competence are less civically engaged, struggle to make economic decisions (e.g., buying a house or taking out loans), and have difficulty determining the risks and benefits of medical treatments (Ancker & Kaufman, 2007; Codding et al., 2017; Crowe, 2010). Without intervention, difficulties with whole number operations are likely to persist over time (Duncan et al., 2007; Morgan et al., 2009). Given that lack of proficiency with basic number operations relates to an increased difficulty throughout education and into adult life, it is no wonder that one of the major findings in the NMAP (2008) report was that competence with whole number operations is a vital skill in mathematics education. The chronic underachievement of students in the U.S. in whole number knowledge coupled with the lasting implications of this underachievement highlight a critical need for students to receive more support to become proficient with basic operations.

Despite the availability of a variety of effective intervention practices to address skill deficits in whole number knowledge (Codding et al., 2011; Gersten et al., 2009), the barriers to implementing such practices are numerous. Some common barriers to implementation include lack of resources, materials, time, training (i.e., acquiring skill necessary), and appropriate application (i.e., when, where, and how to integrate into the curriculum; Bingimlas, 2009; Ertmer,1999; Long et al., 2016; Spectrum K12, 2011). Institutional barriers, such as leadership and support for implementation, funding, and school culture, beliefs, and attitudes, also impact the implementation of intervention practices (Ertmer, 1999; Twyman & Sota, 2016). Implementing an intervention requires that schools use precious resources in order to meet the needs of struggling students. Given that technology-delivered and traditional modes of intervention (e.g., teacher-mediated, self-managed, and peer-mediated) may use different resources, it is important to examine whether the effects of different modes of intervention produce comparable results for students struggling with math computation.

Previous researchers have theorized that the question regarding the efficacy of different modes of intervention delivery is not whether the platform (i.e., teacher-mediated, peer-mediated, self-managed, or technology-mediated) is effective but rather posited that these platforms are merely means of delivering instruction (Clark; 1983, 1985, 1994). Researchers further posited that if the intervention itself includes the evidence-based instructional design principles than the mode of intervention delivery should not account for any unique effects (Twyman & Sota, 2016). However, few studies have empirically examined the idea that the mode of intervention delivery has no impact on student performance beyond the instructional design principles included. If this is in fact true, then the decision of which mode of intervention delivery to adapt within a school setting could largely be made based on the resources of the school thus alleviating common barriers to implementation.

To address the above implementation barriers, educators need a menu of effective practices which promote student whole number operations knowledge that require different resources (e.g., materials, level of personnel support, and instructional applications) so that they could choose the best practice for the unique needs of their students as well as the unique set of resources available to them. Therein providing critical support to students in the U.S. with a documented history of underachievement in whole number operations while accounting for common implementation barriers. In Study 1, this dissertation aimed to examine the instructional design components and whole number operations outcomes of studies which compared two or more modes of intervention delivery. In Study 2, this dissertation tested two class wide interventions with different modes of intervention delivery and a combination intervention on the multiplication fluency of third grade students.

Study 1: Instructional Delivery and Design Principles on Whole Number Operations

Previous meta-analyses have examined the effects of technology-mediated interventions on student mathematics achievement broadly (Burns & Bozeman, 1981; Cheung & Slavin, 2013; Li & Ma, 2010; Moyer-Packenham & Westenskow, 2013; Rosen & Saloman, 2007; Seo & Bryant, 2009; Sokolwski et al., 2015; Steenburgen-Hu & Cooper, 2013; Young, 2017), but none have focused solely on whole number knowledge. Additionally, in some of these studies (e.g., Cheung & Slavin, 2013 & Young, 2017) the type of technology examined was very broad, including supportive technologies such as calculators, interactive whiteboards, and media presentations. Moreover, previous meta-analyses did not exclude studies with business as usual or no treatment control conditions. Therefore, the effects of technology-mediated interventions in previous meta-analyses could be inflated because greater growth is expected when comparing to a no-treatment control than when a treatment is compared to an active control condition (Petersen-Brown et al., 2019). This is problematic because within a multi-tiered systems of support (MTSS) framework, students in need of an intervention would not receive a treatment or continue instruction, which is not meeting their needs. Therefore, an alternative treatment to a technology-mediated intervention is likely a teacher-mediated, peer-mediated, or self-managed intervention.

The importance of using active control conditions has been highlighted by multiple researchers (Gersten et al., 2000; Lewis-Snyder et al., 2002; Petersen-Brown et al., 2019; Pressley & Harris, 1994). To ensure that researchers are drawing appropriate conclusions, it is critical to match the components of the interventions as closely as possible to create rigorous research conditions (Petersen-Brown et al., 2019). Instructional design principles are a set of practices which promote student learning and help educators create, improve, and evaluate instructional practices (Gustafson & Tillman, 1991; Okey, 1991; Twyman & Sota, 2016). This dissertation's first study examined whether instructional design principles within the primary studies were closely matched (in order to ensure that the interventions being used were equivalent) so that the mode of intervention delivery could be evaluated. The purpose of this dissertation's first study was to use meta-analytic procedures to compare the effects of technology-mediated interventions to traditional modes of intervention (teacher-mediated, peer-mediated, and self-managed) on student whole number operations knowledge in studies which compared the two modes of intervention delivery (e.g., technology-mediated and teacher-mediated, peer-mediated, or self-managed). In addition to examining the effects of interventions on student whole number operations knowledge, the first study also sought to examine the quality of the existing literature according to standards used by Jitendra and colleagues (2011).

Study 2: Effects and Generalization of Intervention Modality on Student Math Proficiency

The purpose of this dissertation's second study was to examine the effects of two class wide interventions (e.g., reciprocal peer tutoring and iPad delivered flashcards) and a combination of the two interventions on student multiplication fluency. The combination condition was included to determine if the effects of an intervention using multiple modalities resulted in improved outcomes beyond that of single modality technology-delivered interventions on proximal measures. Previous research supports the use of technology-mediated interventions and reciprocal peer tutoring to support student engagement (Fantuzzo & Ginsburg-Block, 1998; Greenwood, 1991; Haydon et al, 2012); therefore, engagement and off-task behavior of the participants was monitored.

Theoretical Background

Given the pervasiveness of student underachievement in basic number operations, schools should provide class wide interventions that target this foundational skill gap (Vanderheyden et al., 2012). Therefore, educators need effective intervention practices that support student learning while allowing for schools to utilize the resources available. Early theories of learning provide insight into some basic strategies that promote learning while using minimal resources. Two such theories include stimulus-response theory (S-R theory; Burke, 1982) and operant conditioning (Skinner, 1938), which posit that learning occurs when a combination of stimulus, response, and reinforcement are presented. Based on these theories, the theory of programmed learning suggests that when learning academic skills, question, answer, and positive corrective feedback are equivalent to a stimulus, response, and reinforcement (Najarian & Vasilache, 2012). In addition to requiring few resources, the use of carefully constructed practice opportunities (e.g., question and response) and immediate feedback for improving student math performance goes beyond theory and is supported by decades of research (Baker et al., 2002; Burns et al., 2006; Codding et al., 2011; Daly et al., 2007; Fuchs et al., 2008; NMAP, 2008). However, a recent study found that commonly used mathematics curriculums do not provide sufficient opportunities to practice for students to become proficient (Doabler et al., 2012). Thus, providing more evidence supporting the use of a class wide intervention to target student math fact fluency, an area of underachievement for students on national assessments for the past three decades (NCES, 2019).

Evidence Based Interventions

Two approaches which incorporate high rates of feedback and practice that can be embedded into regular classroom routines during core instruction and demonstrate strong research support are technology- and peer-mediated interventions (Bryant et al., 2015; Cates, 2005; Fantuzzo et al., 1992; Okolo, 1992; Rhymer et al., 2000; Tamim et al., 2011). Both these approaches have the potential to provide differentiated instruction and require little support from the classroom teacher to implement. Moreover, research supports both technology-delivered interventions (Bryant et al., 2015; Tamim et al., 2011; Okolo, 1992) and peer tutoring (Fantuzzo et al., 1992; Rhymer et al., 2000) as methods that can increase math computation performance in both fluency and accuracy.

Peer tutoring. Reciprocal peer tutoring is a peer-mediated intervention in which student dyads work together to complete a learning task and each student serves as a tutor and tutee. (Dufrene et al., 2005). Reciprocal peer tutoring has been shown to improve the mathematics achievement of students in 3rd through 6th grade (Bowman-Perrott et al., 2013; Dineen et al., 1977; Fantuzzo et al., 1989; Fantuzzo & Ginsburg-Block, 1998; Kunsch et al., 2007; Maheady et al., 2001; Robinson et al., 2005; Rohrbeck et al., 2003).

In addition, peer tutoring interventions require few materials, time, and training while being flexible to use across skill areas; therein reducing common barriers to implementation. The large quantity of support for the use of reciprocal peer tutoring to improve academic skills across tasks, environments, and students makes it a practical method to individualize instruction with limited impact on the classroom teacher to provide supports (Cates, 2005). In addition to the academic benefits of peer tutoring, studies have found that high rates of engagement are associated with participation in peer tutoring (Fantuzzo & Ginsburg-Block, 1998; Greenwood, 1991). Even further, increased engagement was highly correlated to improvement on a math post-test (Fantuzzo & Ginsburg-Block, 1998; Ginsburg-Block, 1998).

iPad-delivered flashcard drill. Experts have recommended that professionals incorporate technology into mathematics teaching to alleviate implementation barriers, provide individualized instruction, and allow students ample opportunities to practice (Hayden et al., 2012; National Council of Teachers of Mathematics, 2013; NMAP 2008). The use of technology-mediated interventions has been found to have a positive effect on student math performance (Burns et al. 2012; Chueng & Slavin, 2013; Hattie, 2009; Kiru et al., 2018; Li & Ma, 2010; NMAP, 2008; Tamim et al., 2011). One form of technology commonly found in classrooms is the iPad. iPads have a variety of applications (apps) with varying degrees of educational content. One type of app on iPads is flashcard drill practice. Technology-delivered flashcard drill has been found to have small to moderate impacts on student mathematics performance (Duhon et al., 2012; Rich et al., 2012). The use of iPads to deliver such practice provides each student with a means of practicing their basic facts at their own pace and the iPad can give the student immediate corrective feedback regarding their performance. This means that the teacher and their peers are not responsible for providing this critical feedback thereby alleviating common implementation barriers related to providing opportunities to practice and immediate corrective feedback. In addition, studies have demonstrated that students show high engagement while using iPads to practice mathematics (Haydon et al., 2012; McKenna, 2012). This higher engagement has also been associated with improvement in math performance when using an iPad to practice (Evans et al., 2015; O'Malley et al., 2014).

General Purpose of Studies

The persistence of student underachievement in whole number operations coupled with the numerous barriers to implementing evidence-based practices in schools reveals a need to provide a menu of effective evidence-based interventions that require different resources so that schools may choose practices which fit their unique needs. This dissertation sought to address this need by (a) evaluating existing literature which compared two modes of intervention delivery on student whole number operations knowledge and (b) examining the effects of two class wide interventions and a combination of the interventions on student multiplication fluency. Study 1 synthesized 14 existing studies comparing two treatment conditions on student whole number operations knowledge. The results of this meta-analysis suggest that both technology-mediated and traditional methods (i.e., self-managed, teacher-mediated, and peer-mediated) produce meaningful gains in student whole number knowledge. Specifically, there was not a significant difference between modes of intervention delivery on student math performance. Study 2 compared the effects of class wide peer tutoring, iPad delivered flashcards, and a combination of the two methods on student multiplication fluency. The results of this randomized control trial revealed that all treatment groups had significant improvement from pre- to post-test on proximal measures of multiplication fluency. There were few differences between treatment conditions on the measures. One significant treatment comparison was the comparison between the iPad condition and combined condition on the iPad CBM, which favored the iPad condition.

CHAPTER 2

Examining Effects of Instructional Delivery and Design Principles on Whole Number Operations: A Meta-Analytic Review

For over 3 decades, more than 50% of all 4th and 8th grade students have not meet proficiency standards in mathematics on the National Assessment of Educational Performance (NAEP; NCES, 2019). Specifically, national averages on the whole number operations subscale of the NAEP continue to fall below proficiency standards for 4th grade students, meaning that most students in the U.S. are not able to demonstrate whole number knowledge on this assessment (NCES, 2019). In fact, students in the U.S. demonstrate less competence with whole number operations than students from other developed countries, according to the results of the number subtest of the 2015 Trends in International Mathematics and Science Study (TIMSS; NCES, 2015). These data are concerning given that mastery of basic number operations is critical for later success with fractions, algebra, and other advanced mathematics topics (Fuchs et al., 2016; Jordan et al., 2017; NMAP, 2008).

Although existing empirical evidence indicates that educators can select a variety of effective intervention practices to address skill deficits (Codding et al., 2011; Gersten et al., 2009), the barriers to implementing such practices are numerous, including lack of resources, materials, time, and compatibility within the curriculum (Long et al., 2016). Given the documented chronic underachievement with whole number operations and the lack of resources to implement evidence-based practices to address students' needs, technology may be a useful alternative (Chueng & Slavin, 2013; NMAP, 2008). The effects of technology-mediated interventions have yielded small to moderate effects on student math performance (Slavin et al., 2009; Tamim et al., 2011) however little is known about how these options compare to teacher-mediated, peer-mediated, or self-managed interventions. The results of such comparisons are critical to inform resource allocation within schools and to provide information regarding the most efficient and effective ways to support student mastery of basic number operations. The purpose of this study is to provide a meta-analytic review of the existing literature which compares modes of intervention delivery while addressing students' proficiency with whole number operations knowledge.

Whole Number Knowledge

Practice and repetition are critical instructional practices designed to promote mastery with basic facts (Daly et al., 2007; Doabler et al., 2012). For low achieving or young students, more practice may be required to achieve mastery (Burns et al., 2015; Stickney et al., 2012). Students who are not proficient with basic mathematics operations have trouble with advanced mathematics concepts including word problem solving, algebra, and rational numbers (Fuchs et al., 2006; Namkung et al., 2018). For example, a recent longitudinal study found whole number knowledge was a predictor of later fraction knowledge achievement, therefore students with low whole number knowledge were more likely to exhibit difficulty with basic fraction concepts and procedures (Jordan et al., 2017). Whole number knowledge not only affects fraction knowledge but also impacts students' ability to utilize algebraic principals because this skill is dependent on mastery of whole number operations (Ketterlin-Geller & Chard, 2011; Fuchs et al., 2016).

Underachievement in foundational mathematics skills are likely to persist throughout students' educational experiences and into their adult lives without intervention (Duncan et al., 2007; Morgan et al., 2009). Students who do not master whole number operations are less likely to attend college and have lower career success and incomes (Adelman, 2006; Attewell & Domina, 2008; Dowker, 2005). Moreover, adults with low mathematics competence are less civically engaged, struggle to make economic decisions (e.g., buying a house or taking out loans), and struggle to determine the risks and benefits of medical treatments (Ancker & Kaufman, 2007; Codding et al., 2017; Crowe, 2010).

Strong evidence indicates that fluency and accuracy with all four whole number operations is essential for students to access higher level math skills and concepts (Carr et al., 2008; Carr & Alexeev, 2011; Fuchs et al., 2006; Jordan et al., 2013). For students who are identified as benefiting from additional supports with whole number knowledge, the Institute for Education Science (Gersten et al., 2009) recommends that 10-minutes of each intervention session be allocated to practice with basic whole number operations. This recommendation was offered in context of national expert panels recognizing that too few opportunities to practice are provided within existing classroom settings to promote proficiency with whole number knowledge (Burns et al., 2006; Codding et al., 2011; Haring & Eaton, 1978; NMAP, 2008). Consequently, it is important to identify effective intervention options to promote accuracy and fluency with computation skills while considering whether delivery mechanisms may offer schools options that could be more feasible and usable, depending on individual school and teacher needs.

Instructional Design Principles

Fortunately, research supports the use of instructional design principles to promote student learning (Gustafson & Tillman, 1991; Okey, 1991; Twyman & Sota, 2016). Instructional design is a means of making instruction more effective and efficient through the process of designing, developing, and delivering instruction (Gustafson & Tillman, 1991). Instructional design includes both "big" and "little" factors that impact design decision making (Gustafson & Tillman, 1991). Big factors are predetermined by the environment and include large blocks of content and span long periods of instructional time (Gustafson & Tillman, 1991). Whereas little factors are a set of strategies which are more relevant to individual programs and lessons. The present review focused on examining the little factors of instructional design. Some of these instructional design principles include; (a) attention/motivation, (b) recall of prerequisite skills, (c) explicit instruction, (d) opportunities to practice, (e) immediate, corrective feedback, (g) cumulative review (i.e., subdivided skills combined and reviewed), (h) progress monitoring, and (i) retention and transfer of new skills (Gustafson & Tillman, 1991; Okey, 1991; Twyman & Sota, 2016). The identification of such principles provides a framework for not only creating and improving instructional practices but also for evaluating existing practices (Twyman & Sota, 2016).

Implementation Barriers

Despite access to instructional design principles, the implementation of interventions, regardless of mode of intervention delivery, is not without barriers. Recent surveys of educators' practices suggest that implementation of research-based academic interventions is relatively uncommon (Spectrum K12, 2011). The most commonly reported barriers to intervention delivery are related to the intervention itself; notably, the lack of training, time, resources (i.e., intervention materials, time, personnel), and compatibility (e.g., intervention content fit within instruction; Long et al., 2016; Spectrum K12, 2011). Therefore, if interventions with good instructional design principles are equally effective at improving student outcomes regardless of mode of intervention delivery (i.e., teacher-mediated, self-managed, peer-mediated, or technology-mediated), than the decision of which mode of intervention delivery to adapt within a school setting could largely be made based on the needs of the school and the ability to address the above barriers within each school. However, few studies have empirically examined the idea that the mode of intervention delivery does not impact student performance beyond the instructional design principles included.

Modes of Intervention Delivery

Previous researchers have theorized that the question regarding the efficacy of different modes of intervention delivery is not whether the platform (i.e., teacher-mediated, peer-mediated, self-managed, or technology-mediated) is effective but rather that these platforms are merely means of delivering interventions (Clark; 1983, 1985, 1994). Moreover, researchers further posited that if the intervention itself includes the evidence-based instructional design principles discussed above then the platform of intervention delivery should not account for any unique effects (Twyman & Sota, 2016). However, there has not been a synthesis of the literature that has empirically tested whether the mode of intervention delivery has differential effects on student mathematics performance when the instructional design principles of the interventions are held constant.

For the purpose of the present study, four modes of intervention delivery were included, (a) teacher-mediated, (b) peer-mediated, (c) self-managed, and (d) technology-mediated. A simple way to determine the mode of intervention delivery is to examine how the intervention is being delivered. For example, if a student is practicing flashcards with a peer, the peer is delivering the intervention (i.e., prompting the target student to answer the question). Therefore, this is an example of a peer-mediated intervention. In contrast, if the target student is practicing flashcards on their own, the student is prompting themselves to answer each fact (i.e., self-managed intervention). Teacher-mediated interventions can be defined as an intervention which is delivered by a teacher or other adult (e.g., classroom or special education teacher, interventionist, or researcher). For the purpose of the present study, a teacher-mediated intervention does not include class wide instruction because within a response to intervention (RtI) framework, class wide instruction is considered the least intensive form of instruction delivered to students (Mellard et al., 2010). Therefore, class wide instruction, while delivered by a teacher, is not equivalent in instructional intensity to peer-mediated, self-managed, and technology-mediated interventions. For this reason, class wide instruction was not included as a comparison condition in the present study. The last mode of intervention delivery included in the present study is technology-mediated interventions, which are interventions where a technological device (e.g., computer, laptop, or tablet) provides instruction or manages the completion of intervention procedures.

Previous Meta-analyses

While many studies have examined the effects of technology-mediated interventions on academic achievement broadly, fewer have focused solely on mathematics. Nine meta-analyses and one second order meta-analysis were found which examined the effects of technology-mediated interventions on mathematics achievement broadly. These studies found that for students in Kindergarten through 12th grade technology-mediated interventions had a small to moderate positive effect on mathematics achievement. These meta-analyses varied in whether they included only students in general education (e.g., Rosen & Saloman, 2007; Streenburgen-Hu & Cooper, 2013), students with learning disabilities (e.g., Seo & Bryant, 2009), or all students regardless of educational placement (e.g., Burns & Bozeman, 1981; Chueng & Slavin, 2013; Li & Ma, 2010; Moyer-Packenham & Westenskow, 2013; Sokolwski et al., 2015; Young, 2017).

However, the meta-analyses on this topic have multiple limitations. First, they do not provide isolated information on whether technology-mediated interventions are effective for specific mathematics domains (e.g., calculation, word problem solving, algebra, fraction knowledge, etc.). Second, in some of the studies (e.g., Cheung & Slavin, 2013 & Young, 2017) the technology included was very broad (e.g., calculators, interactive whiteboards, multimedia presentations, etc.). Such technologies have fundamentally different uses and are frequently used in conjunction with other modes of intervention delivery (e.g., using a calculator during self-managed practice). The inclusion of this broad definition further confounds interpretation of effectiveness of these tools because these technologies are not the mode of delivering instructional principles but may also serve as a means to supplement or support practice through traditional modes of intervention delivery such as teacher-mediated, peer-mediated, or self-managed interventions. Finally, none of the previous meta-analyses excluded studies with business as usual or no treatment control conditions. This means that the effect sizes could be larger than expected because when comparing a treatment to no-treatment the effect size is larger than when two treatments are compared (Petersen-Brown et al., 2019). Moreover, for students in schools, the alternative to a technology-mediated intervention is likely a teacher-mediated, peer-mediated, or self-managed intervention rather than no intervention, or continuing with instruction that is not meeting students' needs. Therefore, the results of these studies have little utility for practitioners.

The present review expands the current literature in the following ways. First, only studies which examined whole number math computation were included. Second, the technology used to deliver the interventions was limited to computers and tablets. Lastly, the included studies had an active control condition, meaning that the technology-mediated intervention had to be compared to a traditional mode of intervention delivery (i.e., teacher-mediated, peer-mediated, or self-managed) which had equivalent instructional design principles, dose, and skill level. Previous meta-analyses examining the effects of technology-mediated interventions on student mathematics achievement did not restrict the comparison conditions to only active treatment controls. Researchers have suggested that when comparing interventions, it is critical to ensure that the interventions are as similar as possible to create rigorous research conditions which reduce the likelihood of drawing inappropriate conclusions (Gersten et al., 2000; Lewis-Snyder et al., 2002; Petersen-Brown et al., 2019; Pressley & Harris, 1994). Ensuring that the interventions are closely matched (e.g., instructional design principles, dose, and target skill) enables researchers to examine whether the mode of intervention delivery (technology-mediated, teacher-mediated, peer-mediated, or self-managed) is effective at improving student performance.

In addition to examining the effects of interventions on student whole number operations knowledge, the present study also sought to examine the quality of the existing literature. When conducting a review of the literature, it is important to consider the quality of the studies included in the analyses to ensure that claims you are making are appropriate given the quality of the research used to substantiate those claims (Cooper et al., 2009). Jitendra and colleagues (2011) utilized the quality standards of Gersten and colleagues (2005) for group design studies to evaluate the quality of articles included in their meta-analysis. These quality standards are widely used and account for differences in study design type to ensure quality between group and single-case design studies (Jitendra et al., 2011).

Purpose

Due to the profound underachievement of students in whole number operations coupled with the negative impacts of this underachievement on future success (e.g., difficulty with advanced mathematics and adult daily living) it is imperative to provide educators with an array of effective and efficient options to support student whole number operations knowledge. While research indicates that technology-mediated instruction may have positive effects on student mathematics outcomes, little is known about how these outcomes compare to traditional modes of intervention delivery. The comparison between technology-mediated and traditional modes of intervention would provide critical information regarding what works for whom and under what conditions. Additionally, it would help to inform instructional decision making so that school professionals can make informed decisions on how to best use their limited resources. The purpose of the present study was to use meta-analytic procedures to compare the effects of technology-mediated interventions to traditional modes of intervention delivery on student whole number operations knowledge.

Based on the purpose of the present study as well as considerations to quality of the literature the research questions were as follows:

(1) What are the features (i.e., design type, intervention used, age of participants, etc.) of the studies comparing modes of intervention delivery?

(2) What is the quality of studies which have compared modes of intervention delivery according to quality standards set forth by Jitendra, Burgess, and Gajria (2011)?

(3) Are technology-mediated interventions as effective as teacher-mediated, peer-mediated, or self-managed interventions on student whole number operations?

Method

Literature Search and Inclusion Criteria

Three search strategies were utilized to collect studies for possible inclusion in the present meta-analysis. The first step in the search process was to consult with a resource librarian to establish search terms, data bases, and limits. This consultation took place at the University of Minnesota library in November of 2018. Data collection included a search of five databases for studies that compared teacher-mediated, peer-mediated, or self-managed intervention to a technology-mediated mode of intervention delivery (i.e., computer or tablet) on the whole number operations knowledge of students in preschool through 12^{th} grade.

The databases *Academic Search Premier, ERIC, Education Source, psycINFO,* and *ProQuest Dissertations and Theses* were each searched using the University of Minnesota library database in early November of 2019 and resulted in 3,153 results. Consistent with previous reviews examining technology-mediated intervention, search dates were restricted to 1980 to 2019; because the technology used in schools became readily available in the 1980s (Chueng & Slaving, 2013; Fletcher-Finn & Gravatt, 1995; Liao, 1998). Journal studies and dissertations were included in the search. Appendix A provides a comprehensive summary of the search terms used. Keywords to describe the (a) target population (e.g., K-12), (b) technology-mediated interventions, (c) traditional mode of intervention, and (d) content area of math were used. The final search strategy employed was citation searching. After full screening and inclusion of studies both the forward (i.e., searching for studies which have cited the included studies since their publication) and backward (i.e., searching the reference sections of included studies for additional eligible studies) citation searches of the 14 included studies and 8 relevant literature reviews were searched. The forward citation searches on these studies was conducted in late November of 2019 using Google scholar's *cited by* function. Each citation was entered into the search engine and the titles and abstracts of each resulting study reported as having cited primary study was screened for inclusion. The backward citation search was also conducted in late November of 2019 by screening the title and if necessary, the abstract of all studies which were cited by the primary studies that met our eligibility criteria (described below).

Upon completion of each search strategy the resulting studies were uploaded to Microsoft Excel software. This software allowed the authors to easily search for studies which were duplicates and screen the uploaded citations. This resulted in the removal of 702 duplicates. Further removal of duplicates was completed as part of the screening process. Figure 1 outlines the identification, screening, eligibility, and inclusion of the literature search. The search yielded a total of 3,256 manuscripts and dissertations; after removing duplicates 2,554 remained and were subsequently reviewed for inclusion in the study.

Eligibility Criteria

The following criteria were used to evaluate the fit for the study in the current review: (a) the study compared a technology intervention to a teacher-mediated, peer-mediated, or self-managed intervention; (b) the study used computer or tablet technology (e.g., not Smartboard, calculator, etc.); (c) the study addressed the correct academic subject (i.e., whole number operations); (d) the participants of the study were students in pre-school to grade 12; (e) the study presented quantitative data (e.g., mean scores, standard deviations, effect sizes, t-tests, f-tests); (f) the study was not a narrative review or meta-analysis; (g) the article was written in the English language; (h) the article was not a duplicate of another included study; and (i) single-case design studies were ABAB design (e.g., not alternating treatments or multiple baseline designs) and included a baseline. This resulted in the exclusion of 2,345 studies based on screening and another 195 were excluded after a full text review. Therefore, 14 studies were included.

Categorization of Studies

Study features. Categories included publication type (i.e., peer reviewed or dissertation), study design (i.e., randomized control trial, quasi-experimental, single-case design), participant features (i.e., number of participants, age/grade, sex, reported race, reported English language learner status, eligibility for free and reduced lunch, and eligibility for special education services), and setting features (i.e., type of setting, type of classroom, and urbanicity). Types of settings could include University lab, schools, clinic, or juvenile detention centers. Types of classrooms could include special education classrooms, general education classrooms, or self-contained rooms. Urbanicity was coded as either urban, suburban, rural, or not reported.

Assessment and Intervention features. Assessments were coded based on type of outcome measure (published or researcher created), math operations included (addition, subtraction, multiplication, division, or multiple operations), reported reliability and validity metrics, and assessment modality (technology-mediated or paper-pencil). Interventions were coded on the basis of mode of intervention delivery (technology-mediated, self-managed, peer-mediated, or teacher-mediated), name of intervention program, time spent engaging with intervention (including total number of sessions, frequency of sessions per week, and minutes per session), feedback (i.e., information provided to learners about their performance meant to positively influence future performance), and group size (i.e., class wide, small group, or 1:1).

Comparison between interventions. Studies were coded for the extent to which feedback, math operations, and instructional design principles were comparable between each intervention condition within the study. This was done by coding the presence of each of the instructional design principles discussed above in addition to whether feedback was provided to participants, and which math operations was targeted.

Quality Indicators

Methodological quality of the included studies was evaluated using the quality indicators (QIs) outlined by Jitendra et al. (2011). These standards have been cited in over 100 studies and are based on well-known standards set forth by Horner and colleagues (2005) and Gersten and colleagues (2005). The QIs include evaluation criteria for both single case and group design studies on a 3-point rating system. A score of 1 is given if the QI is not reported or the indicator is not met. A score of 2 means that the indicator is partially met, and a score of 3 means that the indicator is met. Consistent with Jitendra et al. (2011), a study met criteria and was considered high quality or acceptable if it received a minimum score of 2 on each indicator. Any component that received a score of 1 was deemed unacceptable.

Intervention Effectiveness

Information regarding the intervention effectiveness within the included studies was gathered by recording reported effect sizes, quantitative results of the primary dependent variable (i.e., t-test, f-statistic, measures of central tendency), and results of other dependent variables (i.e., social validity data and number of opportunities to respond).

Interrater Agreement

To calculate percent agreement 30% of studies were screened by a first-year doctoral student in school psychology. This graduate student received a 30-minute training session from the first author on conducting inter-rater agreement. In this training, the first step consisted of explaining the inclusionary criteria and the exclusion codes. Next, the coding sheet was explained, along with an operational definition of each code. The rater conducted inclusion and exclusion decisions for 30% of the studies retrieved from the search (after removing duplicates). Any disagreements were resolved as they were encountered between the first author and the rater. Initial inter-rater agreement on the screening of studies was 99.7% for title and abstract screening. Next, the rater screened 102 full-text studies that had been included in the initial screening. Inter-rater agreement on the full-text screening of the studies was 100%. Lastly, the rater coded three studies of the 14 included studies. Inter-rater agreement for the coding of the studies was 97%, 95%, and 93%, respectively. Coding discrepancies were discussed between the rater and first author and were resolved.

In addition to inter-rater agreement, intra-rater agreement was conducted by the first author who coded all studies. Intra-rater agreement is used to ensure that a single rater is consistent in inclusion and exclusion decisions which helps to reduce the chances of coder drift (Cooper et al., 2009). After each phase of inclusion (e.g., title and abstract screening, full-text review, and coding) the first author monitored intra-rater agreement. The first author used a random number generator to select 30% of the studies examined at each phase of inclusion. Intra-rater agreement for the title and abstract screening was 99%. For the full-text review 62 studies were reviewed for intra-rater reliability, 100% agreement was obtained. Lastly, the first author conducted intra-rater reliability on 3 studies after coding. These 3 were randomly selected and intra-rater agreement was equivalent to 100%.

Data Analyses & Effect Size Calculation

For each group design study, means and standard deviations were used to calculate Hedges' g (Hedges, 1981) effect size. When means and standard deviations were unavailable, other information was used to calculate Hedges' g (e.g., p-values, sample sizes, t or F statistics).

The software R (R Core Team, 2016) was used to synthesize the effect sizes into two models. The first model was an intercept only model to calculate the average effect size and heterogeneity across all included effect sizes. The second model was a moderator analysis to evaluate whether certain study factors (i.e., publication year, whether a study had been published or not, and intervention modality; technology-mediated or traditional mode) explained the variance observed. Any effect size outliers that were greater than three standard deviations from the mean were adjusted/down weighted to be equal to three standard deviations from the mean. Consistent with recommendations from Tipton (2014) adjustments for small sample sizes were made using models for robust variance estimation. In addition, studies were weighted based on the inverse standard error (i.e., weighting based on sample sizes; smaller samples are given less weight). The proportion of variance explained (I^2 values) were reported to quantify between-study differences in effect size without sampling error (Borenstein et. al., 2009). The potential of publication bias was evaluated by inspecting funnel plots and Egger's test. The funnel plot was visually examined for skewness or an absence in the bottom left of the plot which would be indicative of publication bias (Cooper et al., 2009). The Egger's test provided additional quantitative data regarding whether publication bias may be present, if the correlation between the effect sizes and their sampling variances is strong this would imply the present of publication bias (Egger et al., 1997). All variables were dummy coded variables with no missing data Descriptive analyses were conducted to provide quantitative information about the various coded features and student outcomes.

Additionally, the overall quality of the literature included in this review was evaluated using the rating scale described above (Jitendra et al., 2011).

Results

Study Features

Of the 14 included studies, five were dissertations or theses and nine were peer-reviewed manuscripts published. The included studies were completed between 1985 and 2017 by 14 unique first authors. Ten of these studies were unique to any prior meta-analysis that examined technology and mathematics. An additional nine single case design studies were not included in these analyses even though they met inclusion criteria. These studies were excluded because, unfortunately, Hedges g can only be calculated for multiple baseline design or phase reversal designs with two or more phases (Hedges et al., 2012, 2013). These studies applied an alternating treatments designs (n = 8) or a reversal design without baseline phase (n = 1). The 14 included studies were quasi-experimental group design studies. Twelve of these were between subject design and the remaining two were mixed methods design studies.

Participants and setting features. Table 1 includes the number of participants across demographic areas as well as the number of primary studies in each setting. The 14 included studies represent a total of 1,054 students (31% male) between pre-school to 8^{th} grade, with most participants in grades 2 through 5 (92.1%). However, four studies (37.2% of participants) did not report sex. Nearly half of the studies (48.9% of participants), did not report race or ethnicity. The remaining studies (e.g., 51.1% of total participants) consisted mostly of white students (31.1%), followed by black students (13.3%), Hispanic/Latinx students (2.3%), Asian students (2.7%), and the remaining

1.7% consisted of participants who reported as Native American (*n* = 10), multiple races (*n* = 6), or other races (*n* = 3). Two studies (*n* = 149 participants) reported whether their participants were English language learners or not, therefore of the total sample of participants (*n* = 1,054), 8.8% were not English language learners, 5.3% of students were English language learners, and 85.9% had unknown English learner status. Three studies reported on free and reduced-price lunch status which included 68 participants who were eligible (6.5% of the total participants). Six studies (*n* = 331 participants) reported including students who were eligible for special education, therefore of the total sample of participants (*n* = 1,054) includes 21.5% of students were not eligible for special education services, 9.9% of students who were eligible for special education services, and 68.6% of students with unknown special education eligibility.

One study was conducted in Canada and the remaining 13 were conducted in the United States. Across studies, settings included 12 elementary schools, 1 middle school, and 1 preschool. Two studies did not report the urbanicity of the setting; those that did, included suburban (*n* = 5), rural (*n* = 4), and urban settings (*n* = 3).

Assessment and Intervention features. Table 2 presents the results of the primary outcome measure of each of the primary studies included in the analyses. The most common assessment tool used across studies was researcher created curriculum-based measures (CBM; *n* = 10), followed by standardized math achievement measures (*n* = 4). Digits correct per one or two minutes (*n* = 6 studies) was the most frequently used dependent variable. All but one study reported that the outcome measure contained single digit facts (*n* = 13). The final study, Campbell and colleagues (1987) utilized a drill program and researcher-designed worksheets which progressed from single-digit to multi-digit operations. Most studies assessed and intervened on a combination of two or more of the four whole number operations (*n* = 6); the least frequently targeted operations were division (*n* = 1) and addition (*n* = 1). The included studies compared at least two different interventions, one technology-mediated intervention and one traditional mode (teacher- student- or peer- mediated) of intervention delivery. Table 3 includes information about the technology-mediated and traditional interventions across the primary studies.

Technology intervention features. Across studies technology used to deliver interventions included computers (*n* = 9), microcomputers (*n* = 3), and tablets (*n* = 1). Six studies included untimed computation tasks (i.e., math problems provided without time constraints) and 6 included timed drill. Eleven studies included eight different commercially or freely available technology programs used to deliver the intervention. In the remaining three studies, the researchers created the technology program to mirror the traditional comparison condition. These researcher-created programs were not available for public use. Rich and colleagues (2017) conducted a replication of Duhon et al. (2012); therefore, the two studies used the same researcher-created program. The programs used in nine studies offered technology-mediated feedback in a variety of ways: (a) visual feedback (e.g., phrases on the screen such as good job, way to go, try again, or check marks, stars, smiley faces), (b) audio feedback (e.g., ring or bell noise), (c) correct response displayed on the screen, (d) speed (e.g., how quickly the student responded), or (e) performance report including score or accuracy. Three studies did not report how the technology program provided feedback and in two studies the technology program did not offer feedback to students. The most frequently used instructional design principles in these studies included drill and practice (*n* = 14) and feedback (*n* = 9). The least frequently used instructional design principles included student verbalizations (*n* = 0), explicit instruction (*n* = 1), cumulative review (*n* = 1), retention and transfer (*n* = 1), and recall of prerequisite skills (*n* = 1).

Traditional intervention features. Across studies the most common mode of intervention delivery was a self-managed task (*n* = 8), followed by peer-mediated practice (*n* = 3), and teacher-mediated practice (*n* = 3). The most common task administered in the traditional interventions was untimed computation task (*n* = 7) followed by timed drill (*n* = 5). Across six different interventions that were delivered, the most used was flashcard practice (*n* = 4). The traditional interventions used in eight studies offered verbal (e.g., good job, try again, stating correct fact) and visual (e.g., placing card in correct/incorrect pile, displaying the correct answer in text) feedback. Four studies did not report whether feedback was given, and the remaining two studies did not offer feedback. The most frequently used instructional design principles in these studies included drill and practice (*n* = 13) and feedback (*n* = 8). The least frequently used instructional design principles included cumulative review (*n* = 0), retention and transfer (*n* = 1), and recall of prerequisite skills (*n* = 1).

Within Study Intervention Equivalence. When examining the instructional design principles between the interventions, all studies used equivalent assessments (i.e., assessments were matched in difficulty and content) between the interventions. However, only five studies included assessments that matched the mode of intervention delivery. In other words, if the student received technology-mediated intervention then the assessment was delivered via technology but if the student received traditional mode of intervention then the assessment was delivered via paper-pencil. Even fewer ($n = 3$) used both cross-modality (i.e., mode of intervention delivery did not match the assessment modality) and matched-modality assessment.

In 13 studies, time (i.e., number of minutes) per session was held consistent between treatment conditions whereas one study used number of opportunities to practice as the measure of consistency between the treatment conditions. Table 4 represents the instructional design principles present both within and between the included studies. Of the 14 included studies, half matched all the instructional design principles applied in each intervention condition. Two studies matched all but one of the instructional design principles between the two interventions and five studies had two or more instructional design principles which were not matched between conditions. Seven studies matched the visual aspects across intervention modalities (i.e., number of facts presented, problem solving strategies [e.g., number line, matrices, other visual supports], and visual feedback e.g., [colors, stars, checkmarks, crossed out attempts, correct answer displayed]).

Quality Indicators

Across the 14 studies (see table 5), the overall mean quality rating was 1.85 out of 3 (*range*, from 1.6 to 2.2). Two indicators that rated consistently high were a*ppropriateness of time of data collection* ($M = 3$, $SD = 0$) and t*echniques linked to research question(s); appropriate for the unit of analysis* ($M = 3$, $SD = 0$). The indicators that rated consistently low were *information on intervention agents* ($M = 1.08$, $SD = .23$), *equivalence across groups* ($M = 1.42$, $SD = .74$), and *description of procedural fidelity* ($M = 1.20$, $SD = .56$). Table 5 summarizes the group design study ratings across four domains outlined by Jitendra and colleagues (2011).

Intervention Effectiveness

Each of the 14 studies included at least two effect sizes (e.g., one for technology-mediated and one for traditional modes of intervention). Five studies included multiple effect sizes within each treatment condition due to having multiple assessments (n = 2) or multiple cohorts (n = 3), therefore 44 effect sizes are included in the quantitative analyses. We estimated the overall within (i.e., pre- to post-test) and between group (i.e., technology-mediated to traditional intervention) effects from the 14 studies. Table 6 includes the effect size comparisons within and between the included studies. When examining the effect sizes, there were no outliers (e.g., 3 standard deviations above or below the mean). The potential of publication bias was evaluated by inspecting funnel plots and Egger's test. Results from the multiple methods were not indicative of publication bias. Nonsignificant results were obtained from the Egger's test for both traditional modes of intervention delivery and technology-mediated intervention effects, z = -0.5534, p = 0.5800 and z = 1.0597, p = 0.2893, respectively. Results from the funnel plot did not indicate any studies were missing from the left side of the funnel plot (see figure 2).

The average treatment effect across studies for the technology-mediated treatments (22 effect sizes) was g = 1.22, 95% CI [0.75, 1.69], p < 0.001. The estimated between□study heterogeneity was T^2= 0.5401 and a moderate amount of the between□study variance (I^2= 84.102) was systematic rather than random error. The average treatment effect in these studies for the traditional modes of intervention delivery (22 effect sizes) was g = 0.978, 95% CI [0.12, 1.83], p = 0.029. The estimated between□study heterogeneity T^2= 1.7284 and a moderate amount of the between□study variance (I^2=

94.655) was systematic rather than random error. However, the moderate value of both I^2 statistics indicate that there may be study level variables which are impacting the variance seen between studies, therefore a moderator analysis was conducted, the results of which are below. Figure 3 includes a forest plot which provides an overview of the results of the included studies in terms of effect size and sample size. The boxes show the effect estimates with the confidence interval of the effect. Smaller boxes represent smaller sample sizes whereas larger boxes represent larger sample sizes. The forest plot indicates that many included studies had small sample sizes but there is heterogeneity in the effect size estimates and sample sizes. Additionally, the diamond on the bottom of the plot indicates the pooled effect of the interventions. The diamond is on the vertical line which indicates that there is no difference between the interventions.

Moderator Analysis

Three moderators were included in the analyses published, year of publication, and intervention modality. Published studies represented 32 of the 44 effect sizes whereas the remaining 12 effect sizes were from unpublished studies. In terms of publication, unpublished studies had lower effect sizes on average than published studies, though the difference was not significant (g= -0.333, 95% CI [-1.781, 1.12], $p = 0.5923$). In terms of publication year, studies published prior to 2000 (i.e., 1985-1999) had lower effect sizes on average than those published between 2000 and 2017, though the difference was not significant (g= -0.527, 95% CI [-1.384, 0.33], $p = 0.1921$). Lastly, in terms of intervention modality, technology-mediated interventions had higher effect sizes on average than traditional modes of intervention delivery, though the difference was not significant (g= 0.118, 95% CI [-0.349, 0.586], $p = 0.5734$).

Discussion

The purpose of the current review was to summarize the status of the literature comparing technology-mediated interventions to traditional (i.e., teacher- student-, and peer-mediated) modes of intervention delivery on whole number operations knowledge using qualitative and quantitative methods. The first aim was to provide an overview of the features (i.e., design type, intervention used, age of participants, instructional design principles, assessment type, etc.) of studies which compared technology-mediated to traditional modes of intervention delivery. The second aim of this study was to evaluate the quality of the literature according to criteria set forth by Jitendra et al. (2011). The third aim of this study was to examine the effectiveness of technology-mediated interventions compared to traditional modes of intervention delivery on whole number operations.

The findings of the present study suggest that both approaches (i.e., technology-mediated and traditional modes) represent viable options for intervention delivery when the goal is to improve whole number basic computation skills. The results of the present study indicate that when the instructional design principles of interventions are held constant, the mode of intervention delivery may not have a significant impact on student outcomes. This supports previous research that has suggested that technology is merely a mechanism for delivering an intervention and does not have additional instructional utility beyond the instructional design principles of the intervention (Clark; 19833, 1985, 1994; Twyman & Sota, 2016). The practical implications of these findings suggest that schools have multiple options available to meet the recommendation to provide at least 10-minutes of computation practice with the goal of promoting fluency with basic facts.

This research synthesis also suggests that, based on the poor reporting of demographics in the included studies, it remains unclear for whom technology-mediated interventions may be most effective. Moreover, the poor quality of the included studies limits the conclusions which can be drawn from this body of literature and the present review. Specifically, the included studies lacked information regarding the participant demographics, equivalence between groups at pre-test, and inclusion of procedural fidelity. These omissions limit the claims that can be made regarding whether technology interventions or traditional modes of intervention promote student whole number operations knowledge. Therefore, more high-quality research studies are needed to provide additional support for the use of technology-mediated interventions as an acceptable and effective alternative to traditional modes of intervention delivery for supporting student whole number math operations.

Study Features

Based on the available demographic data reported across studies, students were generally white and enrolled in second through fifth grade in the United States. Few studies reported whether participants were English learners, receiving special education services, or qualified for free and reduce-priced lunch, making it difficult to discern what populations of students these intervention studies were conducted with. Moreover, half of studies did not include participant race or ethnicity information and 25% of studies did not report participants' sex. This lack of reporting of demographic information has also been reported in similar reviews of the literature (Petersen-Brown et al., 2019). Future research should collect and report all demographic information on participants to further understand who may benefit most from interventions using different modes of intervention delivery.

Intervention Delivery Comparisons

It is important to examine modes of intervention delivery given the numerous barriers to implementing evidence-based interventions such as time, personnel, intervention materials, and training (Long et al., 2016). The present review found no significant difference in student outcomes on whole number operations between interventions delivered via technology and those delivered via traditional modes (i.e., teacher-mediated, self-managed, and peer-mediated interventions). However, the average effect size was slightly higher for technology-mediated interventions ($g = 1.22$) compared to traditional modes ($g = .978$). This finding is consistent with previous reviews demonstrating that technology-mediated interventions yielded higher effect sizes than comparison conditions on math achievement (Burns & Bozeman, 1981; Chueng & Slavin, 2011; Li & Ma, 2010; Moyer-Packenham & Westenskow, 2013; Rosen & Saloman, 2007; Seo & Bryant, 2009; Sokolwski et al., 2015; Streenburgen-Hu & Cooper, 2013; Young, 2017). However, these analyses included studies which compared technology-mediated interventions to no treatment or business as usual control conditions. The present study only compared technology-mediated interventions to a traditional mode of intervention with similar instructional intensity therefore the difference in effect sizes between technology-mediated and traditional modes interventions found in the present study were smaller than those reported in previous studies and not significant. This is consistent with the results from Petersen-Brown and colleagues (2019) who found smaller effect sizes when technology-mediated interventions were compared to an active control condition rather than a no treatment or business as usual control.

This review analyzed the extent to which each included study matched the instructional design principles of the technology-mediated and traditional modes of intervention. It was expected that most instructional design principles provided in the traditional modes would also be represented in the technology-mediated interventions given that the purpose of all the included studies was to compare technology-mediated interventions to a traditional mode of intervention delivery. Thus, in these studies, to adequately compare the intervention modality as an independent variable, other variables such as the instructional design principles, had to be held constant (Duhon et al., 2012; Landeen & Adams, 1988). Half of the included studies matched all instructional design principles of the interventions and an additional two matched all but one of the instructional design principles in the interventions. Many researchers posit that matching these components is sufficient to evaluate the differences in the effects between two interventions (Harris, 1988; Landeen & Adams, 1988).

In addition to matching instructional design principles, it is necessary to hold dose constant across interventions (Pressley & Harris, 1994). Each of the studies included in this review kept the number of sessions per week (i.e., frequency), the number of weeks of the intervention (i.e., duration), and the number of minutes per session the same for each treatment condition. Notably in 13 of the included studies, the number of minutes per session rather than the opportunities to respond (OTRs) was held constant between treatment conditions. Previous research suggests that written and verbal response rates are higher than typed responses (Skinner, Belfiore et al., 1997; Landeen &

Adams, 1988). This is an important consideration because all technology-mediated interventions in this review required typed rather than verbal or written responses. Based on this it is possible that the number of OTRs was lower on average in the technology-mediated interventions than the traditional intervention, however this did not translate to differences in effect.

In addition to instructional design principles and intervention dose, to compare modes of intervention delivery researchers should also consider the assessment modality used. All studies included in this review used assessments that were matched in difficulty and content across intervention conditions. However, few studies included assessments that matched the modality of intervention. Previous research indicates that this may have important implications on the accuracy of the results of technology-mediated interventions. More specifically, results from Duhon and colleagues (2012) and Rich and colleagues (2017) indicated that the effects of technology-mediated interventions do not automatically generalize to paper-pencil assessments in the same way that they do to technology assessments. This finding has important implications for future research such that future studies include assessments which are appropriately matched to the intervention modality used but also for examining the efficacy of technology-mediated interventions.

Overall, within the included studies technology-mediated interventions and traditional modes of intervention delivery have similar characteristics in terms of instructional design principles, task difficulty, and intervention dose. Future research should consider incorporating a blend of both technology-mediated and traditional modes and comparing the effects to traditional and technology-mediated interventions by matching instructional design principles, and treatment dose.

Moderator Analyses

The present study examined year of publication and whether a study was published or not as potential moderators to explain differences in effect sizes between studies. The current review examined whether studies published prior to 2000 (e.g., 1980-1999) and those published after 2000 (e.g., 2000-2017) explained differences in effect sizes across the included studies. It might be expected that the effects of technology-mediated interventions improve over time with advancements in such technologies. Although the results of the present study indicate that studies published prior to 2000 (i.e., 1985-1999) had lower effect sizes on average than those published between 2000 and 2017, the difference was not significant. This finding is consistent with Chueng and Slavin (2013), Kroesenburg and Van Luit (2003), and Liao (1998) who found no significant difference among year of publication for studies which examined the effects of technology-mediated interventions on student mathematics achievement broadly. These findings are inconsistent with Li and Ma (2010) and Kulik and Kulik (1987) who found that studies which examined the effects of technology-mediated interventions on student mathematics achievement improved over time.

In addition, results indicate that within the included studies, higher effect sizes were observed for published studies compared to studies that were not published, though the results were not significant. Prior findings that examined publication bias were mixed; one study's outcomes were consistent with the findings of the present study

(Cheung & Slavin, 2013) whereas another study found that published studies resulted in significantly higher effect sizes (Smith et al., 1980).

Quality Indicators

Within this body of literature there are a few areas of strength, including manipulation of the independent and dependent variables, the timing of data collection, and providing a comprehensive description of assessment procedures. For group design studies to be of high quality or acceptable quality they must have met all but one of the 10 essential quality indicators (Jitendra et al., 2011). None of the group design studies in the review met 9 of the 10 quality indicators meaning that the studies in the present review did not meet criteria to be considered high quality or acceptable quality research.

The methodological limitations found herein were consistent with those found by Chueng and Slavin (2013) who reported that many of the studies evaluating technology-mediated interventions have methodological flaws including limited evidence of equivalence between groups, lack of reporting of pre-test differences across groups, and lack of generalized outcome measures. This is problematic because studies with poor methodologies have been found to report inflated effect sizes compared to those meeting rigorous research quality (Chueng & Slavin, 2013). Moreover, without reporting equivalence between groups on multiple variables at pre-test, it is difficult to discern whether the statistical differences found at post-test are due to lack of equivalence between groups prior to intervention. This could result in researchers drawing incorrect conclusions regarding the effectiveness of the interventions. Additionally, the included studies rarely reported whether the interventions were implemented with fidelity or that fidelity had been measured. This is problematic because it is unclear whether both interventions and each intervention session were administered with the same adherence to the instructional practices which make the intervention effective (Collier-Meek et al., 2013; Sanetti et al., 2011). Overall, the included studies did not meet quality standards which is consistent with previous reviews (e.g., Petersen-Brown et al., 2019). Therefore, results should be considered with caution when comparing the effectiveness of technology-mediated and traditional modes of intervention delivery.

Limitations and Future Directions

The contributions of this review need to be considered along with the limitations. First, the scope of studies included may have been limited by the exclusion criteria. For example, studies needed to be available in English which limits the number of studies from outside of the U.S. Additionally, although theses and dissertations were included in this review, the electronic search only captures the projects that were available through online databases. The sample size within the quantitative analysis was small (n = 14) which limited the power available to conduct further moderator analyses. Moreover, due to the small number of studies that met the inclusion criteria, we were unable to evaluate the effects of each the different mode of traditional intervention delivery (i.e., teacher-mediated, self-managed, and peer-mediated) separate from one another. Lastly, the included studies that did not report equivalence between groups on multiple variables at pre-test, it is difficult to discern whether the statistical differences found at post-test are due to lack of equivalence between groups prior to intervention.

Future research may wish to incorporate study quality, assessment modality, and group size as potential moderators for future analysis. Future research should also explore the response rate of interventions requiring different response topographies (i.e., responding via typing versus verbal or written responding) but also should consider if OTR in technology-mediated interventions are more potent or fewer are required to establish an effect than is observed in traditional modes of intervention delivery.

CHAPTER 3

The Impact of Intervention Modality on Basic Number Operations

Students in the U.S. perform below students from other developed countries in basic math computation according to the 2015 Trends in International Mathematics and Science Study (TIMSS; NCES, 2015). Only 41% of U.S. 4^{th} grade students demonstrated proficiency on the National Assessment of Educational Performance (NAEP) in mathematics (NCES, 2019). In fact, the national average score on the number and operations subtest of the NAEP fell below proficiency; indicating that over half of 4^{th} grade students did not meet basic proficiency in whole number operations (NCES, 2019). Taken together this means that most students in the U.S. are not able to demonstrate competence in mathematics on these assessments and this problem has persisted over time. Given that success in advanced mathematics topics (e.g., rational number knowledge, algebraic principles, and word problem solving) relies on mastery of basic number operations, these outcome data highlight a critical need for students to receive more support to become proficient with foundational math skills (Fuchs et al., 2016; National Mathematics Advisory Panel [NMAP], 2008).

Despite a plethora of evidence-based interventions shown to improve student performance with whole number operations (e.g., Codding et al., 2011; Gersten et al., 2009), survey results indicate that these interventions are not being implemented in schools due to numerous barriers (Long et al., 2016). This lack of implementation coupled with the chronic underachievement of students in the U.S. in mathematics highlights the critical importance of providing a menu of effective class wide interventions. However, few studies have compared evidence-based practices that have equivalent instructional design principles to determine if the practices produce similar student gains. Both technology-mediated and reciprocal peer tutoring interventions have yielded small to moderate effects on student whole number knowledge (Fantuzzo & Ginsburg-Block, 1998; Hattie, 2009; Petersen-Brown et al., 2019); however, little is known about how these options compare to one another. Previous researchers have highlighted the importance of such comparisons to provide critical information regarding how the interventions compare in terms of instructional design principles, impacts on student performance, and resources for implementation (Gersten et al., 2000; Lewis-Snyder et al., 2002; Petersen-Brown et al., 2019; Pressley & Harris, 1994). The purpose of the present study is to compare the effects of three class wide interventions (e.g., a peer-mediated intervention, a technology-mediated intervention, and a combined condition) on student multiplication fact fluency across measures of math achievement delivered via technology and paper-pencil methods.

Importance of Whole Number Knowledge

Math fluency is the automatic recall of basic math facts which is measured by determining whether student responding is rapid and accurate (Gersten & Chard, 1999). From a theoretical perspective, building fluent responding is a prerequisite skill to the retention, maintenance, and generalization of the skill to other academic tasks (Binder, 1996; Haring & Eaton, 1978). The evidence aligns with this theoretical perspective in that proficiency with basic facts is a significant predictor of later mathematics achievement (Price et al., 2013). Moreover, the development of basic fact fluency is associated with positive educational outcomes such as attending college, solving complex mathematics problems (e.g., algebraic and rational number knowledge), and interpreting abstract mathematics principles (Kilpatrick et al., 2001; NMAP, 2008; Patton et al., 1997). The critical importance of the development of basic fact fluency coupled with the pervasiveness of student underachievement in mathematics calls for intervention to take place at a class wide level.

Within a multi-tiered system of support framework (MTSS), prevention is aligned with the delivery of an evidence-based core curriculum and effective instructional practices to all students in order to ensure most students (i.e., about 80%) achieve proficiency (Averill et al., 2011). Unfortunately, commonly used mathematics curricula do not provide sufficient opportunities to practice for students to become proficient (Bryant et al., 2008; Doabler et al., 2012), which may contribute to the well documented underachievement for U.S. students on national assessments for the past three decades (NCES, 2019). Solutions to this problem begin by addressing foundational skill gaps at tier 1 during core instruction. When school level performance falls below 80% proficiency or displays low overall levels of mathematics performance, one of the first steps is to provide class wide interventions that target foundational skill gaps (Vanderheyden et al., 2012). Thus, educators need to supplement core curricula with evidence-based interventions facilitating adequate and explicit practice opportunities that can be applied in the classroom context (Bliss et al., 2008; Chafouleas & Riley-Tillman, 2005; Gersten et al., 2009; Long et al., 2016).

Evidence-Based Interventions

The use of carefully constructed practice opportunities (e.g., question and response) and immediate feedback for improving student math performance is supported by decades of research (Baker et al., 2002; Burns et al., 2006; Codding et al., 2011; Daly et al., 2007; Fuchs et al., 2008; NMAP, 2008). Technology and peer-mediated interventions incorporate high rates of feedback and practice that can be embedded into regular classroom routines during core instruction and demonstrate strong research support (Cates, 2005; Bryant et al., 2015; Tamim et al., 2011; Okolo, 1992 Fantuzzo et al., 1992; Rhymer et al., 2000). Both approaches have the potential to provide differentiated instruction, require few material resources and minimal training is needed to implement them. The alleviation of these barriers coupled with the positive impacts on student performance across skills make these two interventions not only positive for students but also feasible for classroom teachers to implement.

Peer tutoring. Peer tutoring is a peer-mediated intervention in which student dyads work together to complete a learning task (Dufrene et al., 2005). Peer tutoring provides frequent opportunities for students to practice and verbalize learning with their peer (Fuchs & Fuchs 2017; Slavin & Lake, 2008; Slavin et al., 2009). Peer tutoring is a well-known method for improving both academic and interpersonal outcomes of students (e.g., time on task, classroom behavior, and student academic self-efficacy; Fantuzzo et al., 1992; Greenwood et al., 1993; Gersten et al., 2009; Hattie, 2009). Research indicates that peer tutoring may benefit students across a range of demographic characteristics (e.g., those with low socio-economic status, English learners, students eligible for special education) with effect sizes ranging from small to moderate across student populations (Ginsberg-Block et al., 2006; Slavin & Lake, 2008; Slavin et al., 2009; Stenhoff & Lignugaris/Kraft, 2007). In particular reciprocal peer tutoring, in which each student within a dyad serves as both tutor and tutee, has shown great promise in promoting the math achievement of students in grades 3 through 6 (Bowman-Perrott et al., 2013;

Dineen et al., 1977; Fantuzzo et al., 1989; Fantuzzo & Ginsburg-Block, 1998; Kunsch et al., 2007; Maheady et al., 2001; Robinson et al., 2005; Rohrbeck et al., 2003). Training students to implement has been found to be simple and it can be used across academic skills (Cates, 2005; Dufrene et al., 2005). In addition to the academic benefits of peer tutoring, multiple studies have found that students who participated in a peer tutoring intervention had higher rates of engagement than those in a control condition (Fantuzzo & Ginsburg-Block, 1998; Greenwood, 1991). Specifically, some researchers found engagement levels as high as 97.5% for students who participated in reciprocal peer tutoring (Fantuzzo & Ginsburg-Block, 1998). This increased engagement associated with reciprocal peer tutoring was highly correlated to higher mathematics scores at post-test (Fantuzzo & Ginsburg-Block, 1998; Ginsburg-Block, 1998).

iPad-delivered flashcard drill. Technology delivered instruction is now commonplace in the classroom with the growing popularity of touch devices (e.g., tablets and iPads) and Chrome Books (Petersen-Brown et al., 2019). Recent reports have called professionals to incorporate technology into mathematics teaching for students at all grade levels to alleviate implementation barriers, provide individualized targeted instruction, and increase opportunities to practice (Hayden et al., 2012; National Council of Teachers of Mathematics, 2013; NMAP 2008). Technology-mediated flashcard drill, specifically, incorporates simple components such as opportunities to practice and feedback without increasing demands on teachers. Moreover, the use of technology-mediated interventions has been found to have a positive effect on student math performance (Burns et al. 2012; Chueng & Slavin, 2013; Hattie, 2009; Kiru et al., 2018; Li & Ma, 2010; NMAP, 2008; Tamim et al., 2011).

Despite the knowledge that technology-mediated drill has positive impacts on student performance, little is known about how iPad technology compares to traditional methods (e.g., peer-mediated, teacher-mediated, or self-managed interventions). Although some researchers argue that all technology is created equal and as it is merely a delivery mechanism for instructional practices (Clark; 1983, 1985, 1994; Twyman & Sota, 2016), others believe that it is important to examine the effects of different types of technology to ensure the effects on student performance are the same as traditional methods (Duhon et al., 2012; Petersen-Brown et al., 2019). iPads are a commonly used type of technology in schools. For example, Apple reported selling 11.4 million iPads to K-12 schools in 2017 (Meaney, 2017). iPads are less expensive and have a more child friendly interface than other forms of technology such as computers or laptops (Kiger et al., 2012). Moreover, studies have demonstrated that students show high engagement while using iPads to practice mathematics (Haydon et al., 2012; McKenna, 2012). Higher engagement has also been linked to greater improvement in mathematics scores when using an iPad to practice (Evans et al., 2015; O'Malley et al., 2014). The widespread use of iPad technology in schools coupled with the limited research regarding its impact on student performance makes examining the effects of this technology critically important to ensure the success of students.

Comparing Evidence Based Practices

Although the effects of peer tutoring and technology-mediated drill and practice are well researched, few studies have compared the effects of the two modalities of intervention on student computation outcomes. Many researchers have stressed the importance of comparing research-based interventions (Petersen-Brown et al., 2019;

Poncy et al., 2007; Skinner et al., 1995/2002; Tamim, et al., 2011). It is arguably more informative to compare novel practices such as technology-mediated interventions to existing evidence-based practices rather than a no-treatment control condition because this comparison can provide information regarding how new practices perform in an applied setting (Gersten et al., 2000; Lewis-Snyder et al., 2002; Petersen-Brown et al., 2019; Pressley & Harris, 1994). There is also value to determining how a new practice compares to practices known to be beneficial (Kilgus et al., 2016; Petersen-Brown et al., 2019). Researchers have suggested that when comparing interventions, it is critical to ensure that they are as similar as possible to create rigorous research conditions, which reduce the likelihood of drawing inappropriate conclusions (Gersten et al., 2000; Lewis-Snyder et al., 2002; Petersen-Brown et al., 2019; Pressley & Harris, 1994). As stated above, both technology-delivered drill and peer tutoring have similar core components such as the presentation of a stimulus (e.g., math fact), a response (e.g., verbal or typed), and immediate, corrective feedback. To our knowledge, no published studies have compared the effects of reciprocal peer tutoring and technology-mediated drill and practice.

Generalization

When assessing the effectiveness of technology-based interventions, it is critical to evaluate the modality of the assessment used to measure the intervention outcomes (Duhon et al., 2012). Assessment across modalities refers to measuring students' skills in a different modality than the one the intervention is provided in (Duhon et al., 2012). Prior research has indicated that when an intervention is delivered by computer but the progress monitoring tool is completed via paper-pencil, skill transfer is not always observed. For example, a study by Duhon and colleagues (2012) found that second graders receiving a technology-mediated intervention displayed higher subtraction fluency when assessed using a technology-based measure than a paper-based measure. Therefore, the skills learned in the technology-mediated intervention did not as easily generalize to the paper-based assessment. Rich and colleagues (2017) conducted a follow-up study, again targeting the subtraction fact fluency of second graders who participated in one of three interventions (e.g., explicit timing [paper practice], computer-based drill and practice, or 4 days of computer drill and 1 day of explicit timing). Rich and colleagues (2017) replicated the results of Duhon and colleagues (2012) and found that for students in their combined condition, one day a week of paper-based practice (e.g., explicit timing) was sufficient to facilitate generalization of the effects across assessment types. These findings have potential implications for evaluating student outcomes because it suggests that alignment between the intervention modality and assessment tool matters.

Purpose

The present study seeks to expand the existing literature by comparing three class wide interventions (i.e., reciprocal peer tutoring, iPad delivered flashcards, a combination of the two interventions) on student multiplication fluency. The combined condition was included to determine if the effects of an intervention using multiple modalities could be replicated based on the findings of the Rich and colleagues (2017). The present study incorporated two proximal measures of multiplication fact fluency, one delivered via iPad and the other delivered via paper-pencil. Additionally, two distal measures of mathematics achievement were used to examine the impacts of these interventions on student mathematics performance broadly. Because previous research has suggested that both reciprocal peer tutoring and technology-mediated interventions promote student engagement (Greenwood, 1991; Haydon et al, 2012), student academic engagement data was collected to examine if there are differences in engagement between the two selected interventions.

The research questions were as follows:

1. Is there a significant difference between reciprocal peer-tutoring, iPad delivered flashcards, or a combined intervention conditions on proximal measures of student multiplication fluency and distal measures of mathematics operations?

2. Are differences observed in students' multiplication fluency between the iPad delivered and paper-pencil proximal measures?

3. Is there a significant difference in the growth of the three treatment conditions (i.e., reciprocal peer-tutoring, iPad delivered flashcards, or a combined intervention) from pre- to post-test on proximal measures of multiplication fluency and distal measures of mathematics operations?

4. Is there a difference in the engagement and off-task behavior between students using the iPad and those who participated in peer tutoring?

Method

Participants and Setting

Participants were recruited from five third-grade classrooms across two elementary schools within a single suburban school district located in the Midwest U.S. and selected by convenience. The school district served predominantly white (76.3%) students who were not eligible to receive free and reduce-priced lunch (82.8%). Most of the students in the school district met proficiency standards in mathematics on national assessments (61.1%). Upon receiving support from the University's IRB and the participating schools' administrators, the first author invited all third-grade classrooms across the two schools to participate in the present study (1 from school A and 4 from School B). Three classroom teachers agreed to allow their class to participate. Passive parental consent forms were sent home with 87 students and collected by the classroom teachers upon return. Parental consent and student assent were obtained for 76 students. Two participants moved out of the district after the start of the study and two participants consistently received special education services during the intervention sessions and therefore missed many sessions. Therefore, 72 participants make up this study. Based on the results of a statistical sample size calculator (Erdfelder et. al., 1996) for an ANCOVA model with fixed and interaction effects, the desired sample size is 88 participants to achieve power of 80% with an alpha level of .0125 and effect size of $f = 0.4$

Participant Demographics. The participants were 52.1% male and the majority were White (82.2%), with the remaining students identified as Black (9.6%), Asian American (5.5%), Latinx/Hispanic (2.7%), and two or more races (1.4%). Most of the participants did not qualify for free or reduced priced lunch (80.8%), were not learning English as a second language (93.2%) and were not eligible to receive special education services (87.8%). Students were randomly assigned to be in one of three treatment conditions, (a) iPad, (b) peer tutoring, or (c) combined. Students were randomized by assigning each a number and entering the number in a random list generator (random.org). The iPad condition included 24 students, the peer tutoring condition included 23 students, and the combined condition included 25 students.

Instructional Setting. All teachers were white females with masters' degrees in education, falling between the ages of 25 and 30 years of age, and teaching experience ranging between two and six years. The school district did not employ an evidence-based core mathematics curriculum, but rather teachers from each grade level compiled lessons and taught grade-level content that aligned with common core state standards. Chrome books were used for various activities across subjects, including mathematics. Classroom teachers reported that on average, students spent approximately 90% of their instructional time in mathematics using traditional methods (i.e., teacher-directed instruction, worksheets, and other non-technology activities). The students spent the remaining 10% of their instructional time engaged in various programs on their Chromebooks.

Intervention setting. At the time of the study, the first author was a fourth-year doctoral student who carried out all intervention procedures. The 20-min sessions were conducted in the students' classroom during time reserved for mathematics. The classrooms were divided in half such that students using the iPads were on one side of the room and those practicing with their peers were situated on the other side of the room.

Measures

The pre- and post-test measures used for all students consisted of two proximal measures of multiplication fact fluency and two distal measures of math achievement. The proximal measures were three forms of a 2-min paper-pencil curriculum-based measure (CBM) and three forms of a 2-min iPad-delivered CBM. The median score for each student on each assessment was used. The two distal measures were the math fluency and math calculation subtests of the Woodcock-Johnson Fourth Edition, Tests of Achievement (WJ-IV; Schrank et al., 2014). The primary dependent variable in this study was student's scores in digits correct per 2-min (DC2M) on the paper-pencil CBM and the iPad delivered CBM at pretest and posttest. A social validity measure was administered to students during post-test to gather information related to the students' acceptability of the intervention procedures.

Paper-pencil curriculum-based measures (CBMs). Six single skill multiplication CBM probes with numerals from 0 to 12 were obtained from AIMSweb (Shinn, 2004). Three probes were administered at pre-test and the remaining three forms were administered at post-test, with the median score among these three forms at each time point serving as the final score for each participant. The use of multiple probes has been shown to provide a more accurate score and reliability across alternate forms of a CBM (Methe et al., 2015). Standardized administration and scoring instructions were followed (Shinn, 2004). Digits were considered correct if the number was correctly written and was in the correct place value location. Digits were considered incorrect if the number was (a) not in the correct place value location, (b) the incorrect digit, or (c) illegible (Shapiro, 2011). Scoring consisted of counting the number of digits correct to determine the number of DC2M. Each probe consisted of 84 randomized multiplication facts with operands ranging from 0-12. The facts were presented in 6 rows and 7 columns across two pages. Alternate form reliability of the CBM multiplication probes was computed for this study was .818 on average (range, $r = .630$ to .936). According to Foegen and colleagues (2007), CBM reliability coefficients ranged from .80 to .92 and validity coefficients ranged from .40 to .80 across types of CBMs. Criterion validity of single skill CBM measures ranges from 0.74 to 0.83 (Christ et al., 2008).

iPad delivered CBMs. The iPad application (app) used to administer the iPad CBMs was Timed Test (FormSoft Group, Ltd. 2012). The platform enabled the first author to: (a) set up a timed test for multiple users, (b) view the users previous test scores and answers, and (c) set the test time, number of problems, and desired operands included in the test. For the purpose of the present study the first author entered each participant's name in the app and set the test settings to be consistent with the CBM probes, therefore, the timer was set to 2-min and included 84 multiplication facts with operands between 0-12. Students were instructed to select begin and answer as many multiplication problems as possible in 2-min using the on-screen number pad and their finger (Appendix B). If students did not know the answer to a problem, they could skip the problem by swiping from right to left on the screen. The iPad app automatically stopped the students after 2-min. The app provided a summary of each completed assessment which was printed and subsequently scored by the first author using the same scoring procedures that were used for the paper-pencil CBM, described above. The mean alternate form reliability of the iPad delivered CBM computed for this study was .805 (range, r = .681 to .893). On average, concurrent validity between the paper-pencil CBMs and iPad CBMs in this study was .738 (range, r = .547- .868).

Woodcock-Johnson Test of Achievement-IV (WJ-IV; Schrank et al., 2014). Pre-testing and post-testing included the math fluency and calculation subtests of the WJ-IV. The calculation task is an untimed measure consisting of various mathematics problems ranging from simple addition to advanced calculus problems. The fluency subtest is a 3-min timed test including single digit addition, subtraction, and multiplication facts. Standardized administration and scoring procedures were used for both subtests. A students' score was computed as the total number of correct problems. Reliability of the calculation and fluency subtests is .91 and .95, respectively, for children ages 7 to 11 (McGrew et al., 2014). The validity of these mathematics subtests ranges from .68 to .77, for children ages 6-13 (McGrew et al., 2014).

Acceptability survey. The Kid's Intervention Profile (KIP; Eckert et al., 2017) was adapted for this study and used to evaluate student acceptability of intervention conditions. The KIP can be used with any academic intervention and the items are written for easy customization of different intervention procedures (e.g., How much do you like [insert specifics of intervention]; Eckert et al., 2017). Each item included a 5-point Likert scale with the response options; not at all, a little bit, some, a lot, and very, very much. Psychometric data was conducted for writing and includes alternate form reliability (r = .82 to .95) and criterion-related validity (r = .40 to .66). Although the psychometric properties are not available for a math version of this measure, it is anticipated that given the flexibility of the measure there is not a substantial risk to the reliability or validity of the measure. To score the KIP, the sum of the eight items is calculated, where responses of "not at all" are equal to 1 and responses of "very, very much" are equal to 5. For two items (3 & 8) the opposite is true, meaning that for these items a response of "not at all" is equal to 5 and "very, very much" is equal to 1. The scores range from 8-40 with a score of 24 or higher indicating that the student is accepting of the intervention (Eckert et al., 2017).

Systematic Direct Observation. Students were observed using the Behavioral Observation of Students in Schools (BOSS; Shapiro, 1996) to examine whether there was a difference between interventions in terms of student academic engaged time or off-task behavior. The BOSS uses both partial interval and momentary time sampling to categorize students' behavior into engaged time (e.g., passive engaged time and active engaged time) or off task behavior (e.g., off-task motor, off-task verbal, and off-task passive). Active engaged time was defined as time spent engaging with the intervention appropriately (e.g., peer tutors were practicing flashcards with one tutor and one tutee or the students using the iPad were answering facts when prompted). Passive engaged time was defined as time spent engaging in the procedures which were not directly related to solving a math fact (e.g., waiting for partner to count the number correct, shuffle cards, or reviewing summary of practice session on the iPad). Off-task motor was defined as any movement from the designated practice spot (e.g., going to bathroom, sharpening pencil, wandering around the room), off-task verbal was defined as any talking that was not directly related to the task of solving math facts (e.g., chatting with partner, talking to another student), and off-task passive was defined as any behavior that was not related to the intervention procedures (e.g., staring into space, changing settings on the iPad, drawing on the whiteboard, etc.). The first author used a random list generator to identify a pair of students within each class to conduct an observation during each session. This resulted in all dyads being measured at least one time and multiple groups having two observations. Each observation session was coded as either the student engaging in peer tutoring or iPad delivered flashcards, meaning that the students in the combined group were separated into whichever intervention they had been engaged with at the time of the observation.

Procedures

Pre-testing. Pre-testing occurred on two different school days in each participants' classroom. Paper-pencil CBMs and the iPad CBMs were delivered to the participants by the first author. Each student was randomly assigned to either complete the three iPad CBMs before the paper-pencil CBMs or vise versa to control for order effects. Both the iPad and paper-pencil CBM probes were administered in a single session lasting 25-min (i.e., 2-min for each probe plus time for instructions and a 3-min stretch break after completing three probes). The classroom teachers assisted the first author and monitored students during the assessment. The second day of pre-testing was reserved for the WJ-IV fluency and calculation subtests. The fluency subtest was administered first, and the remaining time was reserved for the students to work on the calculation subtest.

Training. The day prior to commencement of the intervention procedures, the first author worked with each participant in small groups of 7-8 students to teach them the appropriate intervention procedures (iPad, peer tutoring, or both). On the day of the training, the first author reviewed expectations and intervention procedures for 5-min. During this time, the first author introduced the materials to the students and instructed them on how to use each appropriately. Then, the first author spent 5-min demonstrating the intervention procedures with a student to the rest of the group. The last 5-min of the training session was reserved for each student to practice the intervention procedures with feedback from the first author.

Intervention conditions. The intervention components and active ingredients for the two intervention procedures is available in table 7. Intervention sessions occurred for five consecutive weeks with four sessions per week. Sessions lasted 20-min which included 5-mins of introduction and preparation followed by 10-min of practice and ending with 5-min of clean up. During the 10-min intervention sessions, students practiced basic multiplication facts using their designated procedures (i.e., iPad or peer tutoring). To reward appropriate behavior, such as respecting others and property, students earned up to 5 stickers during each intervention session that were recorded on a chart (Begeny, 2009). Each student who earned 18 of 20 stickers for the week earned a small prize, such as a pencil, pen, eraser, small toy, or class game time. One teacher agreed only to applying the stickers to the student chart as the reward for meeting behavior expectations.

Reciprocal Peer tutoring condition. Using procedures consistent with previous research, the peer tutoring condition consisted of six components: (a) tutor presents each card to the tutee, (b) the tutee writes the answer on a small whiteboard and shows it to the tutor, (c) the tutor gives corrective feedback and praise to tutee, (d) steps a-c continue until all have been presented and pair continues until timer sounds, (e) the cards are shuffled and the pair switches roles (i.e., tutor becomes tutee), (f) steps a-d are repeated with new tutor/tutee (Fantuzzo et al.,1998). At the beginning of each session, the first author set a timer for 5-min and peer pairs began practice with one student acting as the tutor and one acting as the tutee. With each presentation of a fact by the tutor, the tutee would write a response on their whiteboard. If the tutee could not answer in 3-sec or did not get the answer correct, the tutor provided immediate, corrective feedback (e.g., nice try, 2x5 is 10). Then the tutor would place the incorrectly answered fact on a mat constructed of paper and laminated in the box labeled as “what I am working on”. If the tutee answered a fact correctly in 3-s or less, the tutor provided corrective feedback and praise (e.g., “great job, 3x3 is 9”). Then the tutor would place the correctly answered fact card in the “what I know” box on the mat (Appendix C).

iPad condition. The *Mental Math Cards Challenge* app (McNamara, 2017) was selected by the first author with the following considerations: (a) ease of use, (b) cost, (c) immediate, corrective feedback included, and (d) the drill and practice procedures used mirrored those of the peer tutoring condition. The difficulty of the application was defaulted to easy (which contained facts with operands 0 to 12), the number of practice questions per deck was set to 40 (the highest option), and auto advance on incorrect answers was turned off (this allowed for corrective feedback to be provided). When the interventionist instructed students that the 10-min timer had started, students selected “multiplication” using the touch screen and then selected “practice”. The apps flashcard interface (appendix D) featured a basic multiplication fact in the center of the screen. Below the fact was a number pad where the students could type their answer. In the top right corner was a running total of the number of problems the student had answered correctly and incorrectly. The left corner displayed the question number out of 40, indicating the student’s progress. If the student answered a question correctly, a green check mark appeared over the fact and the word “correct” appeared beneath the answer, then the app proceeded to the next fact. If the student answered the problem incorrectly, a red “x” appeared over the fact and below the fact with the words, “The answer is […]” were displayed. In order to move to the next fact, the student pressed “next”. After the student completed the first 40 problems, the app provided “practice statistics” (Appendix E) which was a summary of the following information: (a) number of questions completed, (b) number of correct answers, (c) number of incorrect answers, and (d) accuracy.

Combined condition. Consistent with procedures from Rich and colleagues (2017) students in the combined condition participated in the iPad condition three days per week. On the fourth day of the week, the pair followed the procedures for the peer tutoring condition.

Post-test. Post-testing occurred over two days and consisted of the same procedures as pre-test. In addition, the students completed the acceptability survey at the end of the first day. The classroom teachers also completed a two-question survey each item included a Likert scale from 1-5 with the anchors being not very likely (1), somewhat likely (3), and very likely (5). The first item asked how likely the teacher was to use technology-delivered flashcards in their classroom in the future and the second question asked how likely they were to use reciprocal peer-tutoring.

Inter-scorer Agreement & Procedural Fidelity

Inter-scorer agreement (ISA) was assessed for 20% of sessions, ISA was calculated by summing the total number of digits attempted and dividing by the total digits agreed upon and multiplying by 100. ISA was completed by a fourth-year school psychology doctoral student, a brief introduction to scoring procedures was provided by the first author prior to completing ISA. The average ISA on the paper-pencil CBM of 100%, (range, 86%-100%). Average ISA for the iPad CBM was also 100% (range, 93%-100%). Lastly, the average ISA for the WJ-IV subtests was 99% (range, 86%-100%).

Procedural fidelity was assessed for both interventionist procedures and student intervention procedures. Fidelity of interventionist procedures was assessed by the classroom teachers for 20% of sessions. Teachers and interventionists discussed intervention procedures prior to implementation and a 6-item checklist detailing the procedural steps. Fidelity of student procedures was assessed by the first author using a checklist detailing the procedural steps of each intervention. One pair of students was randomly assigned for a fidelity check prior to the start of the intervention sessions, and one pair of students were assessed in each classroom on 10 intervention days. Procedural fidelity was calculated by summing the number of correctly implemented steps, divided by the total number of steps on the implementation checklist, and multiplying by 100. Treatment adherence for the interventionist was 100% (range, 83%-100%). Treatment adherence was as follows for students in the: (a) iPad condition 99% (range, 67%-100%), (b) peer tutoring condition 99% (range, 92%-100%) and (c) combined condition 99% (range, 92%-100%).

Research Design & Analysis

A pretest-post-test experimental design with random assignment across three conditions was used. Descriptive statistics of the four measures including means, standard deviations, and correlations were calculated. Additionally, descriptive statistics including an one-way Analysis of Variance (ANOVA) and independent samples t-tests were used to examine the presence of (a) pre-test differences between groups on each of the measures, (b) order effects between participants who were tested on the iPad first versus those who used paper-pencil first at both pre- and post-test on the CBM measures, and (c) differences in DC2M between the two CBM measures within participants at pre- and post-test. Additionally, the rate of improvement (ROI) of each group on the two proximal CBM measures was calculated by subtracting the pre-test scores of each group from the post-test scores and then dividing by the number of weeks of intervention (i.e., 5) to get an average change in DC2M per week for each group on the measure.

In order to determine difference between treatment conditions, four one-way between-groups ANCOVA models (i.e., one for each measure) were conducted. The independent variable in each model was the treatment condition (i.e., iPad, peer-tutoring, or combined) and the dependent variable was the post-test scores on each measure. Participants' scores at pre-test on each respective measure were used as the covariate in this analysis. The assumptions of the chosen inferential statistic, Analysis of Covariance (ANCOVA), were checked. Due to the multiple comparisons being conducted (e.g., four ANCOVA models for each measure) the Benjamini-Hochberg (B-H) procedure was used. This method was chosen due to its utility in controlling for false discovery rates and power in limiting type 1 errors (Williams et. al., 1999). Additionally, the B-H procedure allows the user to sequentially compare the observed p-values for each dependent variable, based on their importance or expected level of effect (Thissen et. al., 2002). Therefore, the B-H corrected level of significance for each dependent variable is as follows; iPad CBM (p= .025), paper-pencil CBM (p= .01875), WJ-IV fluency subtest (p= .0125), and WJ-IV calculation subtest (p= .00625).

In order to evaluate growth on proximal and distal measures from pretest to post test Cohen's *d* effect size was calculated for each group. Consistent with Cohen (1962) small, moderate, and large Cohen's d effect sizes are equal to .2, .5, and .8, respectively. In order to evaluate differences in engagement and off-task behavior between groups, paired samples t-tests were conducted. To do this each observation session was categorized as either taking place during the iPad practice or peer-tutoring practice.

Therefore, the combined group was categorized into the type of practice the student was engaged in at the time of the intervention. Lastly, descriptive analyses were used to evaluation social validity of the intervention procedures and student engagement.

Results

Seventy-two third grade students participated in the present study. The loss of four students mid-way through the intervention resulted in unequal groups across conditions. The peer tutoring condition had 23 students, the iPad condition had 24 students, and the combined group had 25 students. The descriptive statistics are described first for both the proximal and distal measures followed by the inferential statistics. Following these primary analyses, the outcomes for student engagement, off-task behavior, and intervention acceptability are described.

Descriptive Statistics

Table 8 provides the mean score, standard deviation, ROI, and Cohen's *d* effect size for each treatment group across the four measures. There was a strong correlation between the paper-pencil CBM and iPad CBM (r= .872) and a moderate correlation between the paper-pencil CBM and the WJ-IV fluency subtest (r= .566) and WJ-IV calculation subtest (r= .503). There was a strong correlation between the iPad CBM and the WJ-IV fluency subtest (r= .644) and a moderate correlation between the iPad CBM and the WJ-IV calculation subtest (r= .538). Lastly, there was a strong correlation between the WJ-IV fluency and calculation subtests (r= .657).

The results of an one-way ANOVA indicate that the three treatment conditions were not significantly different at pre-test on the iPad CBM ($F(2,67)$= 1.791, p= .175), paper-pencil CBM (*F(2,66)*= .903, *p*= .410), WJ-IV fluency subtest (*F(2,67)*= .253, *p*= .778), or WJ-IV calculation subtest (*F(2,68)*= .767, *p*= .469).

Results of an independent samples t-test indicate there were no significant differences between students who were administered the iPad CBM first and those who received the paper-pencil CBM first during pre-test or post-test sessions. Students who received the iPad CBM first at pre-test did not have significantly different scores than those who received it second (*t(58.55)*= .202, *p*= .841). The same is true for the paper-pencil CBM at pre-test (*t(60.31)*= .502, *p*= .618), the iPad CBM at post-test (*t(49.87)*= -.128, *p*= .899), and the paper-pencil CBM at post-test (*t(47.6)*= -.020, *p*= .984).

A paired samples t-test was used to examine differences in scores on the two CBM measures at both pre- and post-test. Results indicate that participants' scores at pre-test on both the iPad and paper-pencil CBM were significantly different (*t(67)*= 6.533, *p*= .000). Additionally, the participants' scores at post-test were significantly different between the iPad and paper-pencil CBM (*t(67)*= 9.878, *p*= .000). These results indicate that at both pre- and post-test the participants had higher rates of DC2M on the paper-pencil CBM than the iPad CBM.

On the paper-pencil CBM the ROIs for iPad condition (4.56 DC2M per week), peer tutoring condition (4.01 DC2M per week), and combined condition (5.16 DC2M per week) surpassed the expected ROI on this measure for students receiving no intervention. According to Pearson (2011) the expected ROI for 3rd grade students on this measure is a mean of .72 DC2M per week. Large ROIs were also found for the iPad CBM for the iPad condition (5.53 DC2M per week), peer tutoring condition (3.40 DC2M per week), and combined condition (3.21 DC2M per week).

Inferential Statistics

The assumptions of an ANCOVA include normality, linearity, homogeneity of variances, and homogeneity of regression slopes. Preliminary checks were conducted to ensure there was no violation of each of the above assumptions. Normality was evaluated through visual analysis of the Q-Q plots of each measure were not indicative of a violation to the assumption of normality. The relationship the dependent variable and the covariate were best fit through a linear model for each measure. The homogeneity of variances assumption was tested through a Levene's test and for each measure the test was not significant and therefore was not indicative of violations to the homogeneity of variance assumption. The homogeneity of regression slopes was tested by examining the significance of the interaction between the independent variable and covariate, none of these interactions were significant indicating that the homogeneity of regression slopes assumption was not violated.

Paper-pencil CBMs. The results of an ANCOVA signify that the main effect for treatment condition was not significant; $F(2, 66) = .635$, $p = .533$, $\eta_p^2 = .020$, when controlling for pre-test scores. When controlling for treatment condition, the covariate (pre-test scores) significantly impacts post-test scores $F(1, 66) = 102.664$, $p = .000$, $\eta_p^2 = .623$, indicating that pre-test scores account for 62.3% of the variance in post-test scores. Cohen's d effect sizes from pre- to post-test on the iPad CBM were large for the iPad (d = 1.02), peer tutoring (d = 1.03), and combined (d = 1.05) conditions.

iPad CBMs. The results of an ANCOVA indicate that the main effect for treatment condition was significant at the B-H corrected alpha level of .025; $F(2, 65) = 3.902$, $p = .025$, $\eta_p^2 = .113$. Post-hoc comparisons using the B-H corrected alpha level of

.025 indicate that there was a significant difference between the iPad condition and the combined condition on the iPad CBM (*M difference* = 11.925, *SE*= 4.653, *p*= .013), favoring the iPad condition. There was not a significant difference between the iPad condition and peer tutoring condition (*M difference*= 10.486, *SE*= 4.776, *p*= .032) when using the B-H corrected alpha level or the peer tutoring condition and the combined condition (*M difference*= 1.440, *SE*= 4.831, *p*= .767). When controlling for treatment condition, the covariate (pre-test scores) significantly impacts post-test scores $F(1, 65)$ = 82.350, p = .000, ηp^2= .574, indicating that pre-test scores account for 57.4% of the variance in post-test scores. Cohen's *d* effect sizes from pre- to post-test on the iPad CBM were large for the iPad (*d* = 1.30) and peer tutoring conditions (*d* = 1.01) and moderate for the combined (*d* = .72) condition.

WJ-IV Fluency Subtest. The results of an ANCOVA indicate that the main effect for treatment was not significant, $F(2, 66)$= .466, p = .630, partial η^2= .015. When controlling for treatment condition, the covariate (pre-test scores) significantly impacts post-test scores $F(1, 66)$= 87.651, p = .000, partial η^2= .586, indicating that pre-test scores account for 58.6% of the variance in post-test scores. Cohen's *d* effect sizes from pre- to post-test on the WJ-IV fluency subtest were small for the iPad (*d* = .27), peer tutoring (*d* = .21), and combined (*d* = .19) conditions.

WJ-IV Calculation Subtest. The results of an ANCOVA indicate that the main effect for treatment was not significant, $F(2, 67)$= 1.955, p = .150, partial η^2= .058. When controlling for treatment condition, the covariate (pre-test scores) is below the uncorrected alpha level of 0.05 however the B-H corrected alpha level of 0.00625 indicates that the covariate did not significantly impact post-test scores $F(1, 67)$= 77.327,

p = .014, partial η^2 = .092. Cohen's *d* effect sizes from pre- to post-test on the WJ-IV calculation subtest were small for the iPad (*d* = .16), peer tutoring (*d* = .24), and combined (*d* = .41) conditions.

Systematic Direct Observation

On average, students who practiced using peer tutoring had significantly higher rates of engaged time (*M*= 95.67%, *SD*= 9.18) than those practicing with the iPad (*M*= 87.38%, *SD*= 18.43; t(72.590) = -2.699, *p*= .009). Moreover, the students who practiced using peer tutoring had significantly lower rates of off-task behavior (*M*= 10.64%, *SD*= 14.35) than those practicing with the iPad (*M*= 27.91%, *SD*= 27.98; (*t*(73.59)= 3.68; $p <$ .001).

Intervention Acceptability

Students in all three treatment conditions reported high acceptability ratings. No significant differences were observed between conditions; albeit the iPad condition had the highest mean acceptability (31.61; *SD*= 6.243), followed by the peer tutoring condition (*M* = 29.76; *SD*= 7.035) and then the combined condition with a (*M* =29.22; *SD*= 7.154). All three teachers indicated that they were highly likely to use both peer tutoring and technology-delivered flashcards in the future.

Discussion

The present study sought to address existing gaps in the literature by examining the effects of iPad delivered flashcards, peer tutoring, and a combined intervention condition on proximal measures of multiplication fluency delivered via paper-pencil and iPad. No prior published studies, to our knowledge, have made this comparison. Comparing the technology-delivered intervention to the peer-delivered intervention offered the opportunity to also evaluate the impact on cross-modality outcome measures. Lastly, previous research has found that both reciprocal peer tutoring and technology-mediated interventions increase student engagement over control conditions (Greenwood, 1991; Haydon et al, 2012).

The results of the present study indicate that on the paper-pencil CBM, WJ-IV fluency subtest, and WJ-IV calculation subtest there were no significant differences between treatment conditions. On the iPad CBM the intervention condition was significant indicating that students in the iPad condition outperformed students in the combined condition. No other treatment comparison was significant. Students in all conditions made significant gains from pre-test to post-test on proximal measures. Additionally, Cohen's *d* effect sizes were large on both CBM measures and small to moderate on the WJ measures. The ROI statistics for the paper-pencil CBM illustrated gains for students in all three groups that surpassed the expected ROI on the measure for third grade students not receiving intervention (Pearson, 2011). Overall, students yielded higher scores on the paper-pencil CBM measure compared to the iPad measure and students were more academically engaged and less off-task when they were in the peer-tutoring condition as compared to the iPad condition.

Curriculum Based Measures (CBMs)

Students across all conditions had higher rates of DC2M when completing the paper-pencil CBM than the iPad CBM at both pre- and post-test. This may be because response times are slower when typing than when writing (MacArthur & Graham, 1987; Skinner et al., 1997). This could also be because the students have more experience with writing than typing; while their classrooms utilized technology, the time spent using it was much lower than time spent completing traditional activities (e.g., worksheets, paper-pencil tests) that involved writing. The results of the present study support previous research findings that student response times are slower when typing than writing given that the iPad condition performed significantly higher than the combined condition on the iPad CBM and also yielded the highest ROI compared to all conditions. It is possible that typing speed improved for these students as a result of repeated practice using the iPad. Another explanation may be that the iPad CBM was the most proximal measure to the iPad intervention, given the matched modality. This would be consistent with results from Duhon and colleagues (2012) and Rich and colleagues (2017) who found that the technology-mediated intervention yielded the highest score in DC2M on the technology-delivered CBM. It is interesting that the iPad condition outperformed the combined condition but not the peer tutoring condition in the present study; however, a lack of power to detect differences between groups may have been the reason that no significant difference was found between the peer tutoring condition and the iPad condition on this measure.

On the paper-pencil CBM, the differences in effect size and ROI were negligible and there were not significant differences between groups on this measure. This is inconsistent with the hypothesis that the peer tutoring condition would have the greatest growth on the paper-pencil measure. Previous studies found that the paper-pencil condition outperformed the technology condition and combined condition on the proximal paper-pencil measure (Duhon et al., 2012; Rich et al., 2017); however, this was not confirmed in the present study. It is possible that because the time to practice was held constant at 10-min that the potency of opportunities to practice in the iPad condition were higher than that of the peer tutoring condition where students spent 5-min as the tutee and 5-min as the tutor. Previous research suggests that acting as both the tutor and tutee is beneficial for students (Cohen et al., 1982; Franca et al., 1990; Maheady & Harper, 1987); however, it is unclear if it is equally beneficial to be in either role. This, coupled with the idea that the combined group had the opportunity to generalize their learning from the iPad to paper-pencil during the day of peer tutoring practice, may have contributed to the slightly greater effect size observed with the combined group as compared to that of both the peer tutoring and iPad groups on the paper-pencil CBM (Rich et al., 2017). Moreover, while generalization is commonly thought of as an independent step of the instructional hierarchy (Haring & Eaton, 1978), some researchers have argued that generalization should be considered at each stage of skill development (Skinner & Daly, 2010). Regardless, researchers seem to agree that generalization is something that needs to be explicitly included in interventions (Codding & Poncy, 2010; Haring & Eaton, 1978; Stokes & Baer, 1977). The results of the present study, Duhon and colleagues (2012), and Rich and colleagues (2017) support the idea that generalization should be considered when examining the effects of interventions across modalities because the effects of single modality assessments do not necessarily generalize to an assessment of a different modality.

All groups had significant growth from pre-test to post-test on the CBM measures. These results suggest that all three interventions were effective for improving the multiplication fact fluency with this sample of third grade students. These findings are consistent with study 1 which found that there was no difference between interventions delivered via technology and those delivered via traditional methods on student math achievement (Kromminga et al., 2020). However, these results are inconsistent with studies which found differences in students' performance, favoring technology-delivered practice, using technology-mediated versus peer-mediated interventions (Cates, 2005). Moreover, for three of the four measures, pre-test was a significant covariate for post-test scores and explained much of the variance in post-test scores. These results are consistent with research which suggest that initial performance is a significant predictor of post-test outcomes (Clarke et al., 2019).

Woodcock Johnson- IV Tests of Achievement Subtests

Despite the large effect sizes noted for students in each condition across the paper-pencil proximal measure, this performance did not generalize to the WJ-IV fluency subtest. This may have been because the WJ-IV fluency subtest has a combination of all four basic operations whereas the proximal measures only included multiplication which was the target operations of the intervention conditions. Moreover, the WJ-IV fluency subtest is progressive in difficulty meaning that it starts with basic addition and subtraction facts before moving to basic multiplication and division. Therefore, it is possible that some of the students did not reach the multiplication items to display their knowledge. Moreover, the participants scored within the above average range of performance at pre-test, this could indicate that the students had little growth to make on the assessment therefore a ceiling effect was present. Similarly, on the WJ-IV calculation subtest, students did not have significant gains from pre-test to post-test and the effect sizes were small. This may be because the WJ-IV calculation was a more distal measure because the measure contains various mathematics problems ranging from simple addition to advanced calculus problems, only the combined condition yielded significant improvement from pre- to post-test. Perhaps this indicates that practicing across multiple modalities better facilitates generalization of multiple skills compared to practicing on a single modality, as demonstrated by Rich and colleagues (2017). Moreover, the support for the use of multiple modalities is present in literature which has found that computer assisted instruction is more effective as a supplement to traditional instruction rather than a replacement (Burns et al., 2012; Cheung & Slavin, 2013; Fuchs et al., 2006; Musti-Rao & Plati, 2015). However, this difference was not significant in the between condition analyses, which may be due to the small sample size. A moderate effect size difference was found between the iPad and combined conditions and the peer tutoring and combined conditions, favoring the combined condition. Future researcher should consider examining whether there is a benefit to using multiple modality interventions to support student progress on distal measures of achievement.

Systematic Direct Observation

Students participating in the peer tutoring intervention were significantly more engaged in the intervention than those using the iPad and students using the iPad were significantly more off-task than those participating in peer tutoring. These results are inconsistent with a study by Bryant and colleagues (2015) who found that students using the technology-delivered intervention were more engaged than those in the teacher-delivered intervention condition, however, the teacher-delivered intervention was a combination of small group and independent worksheet practice therefore it is possible that this form of practice is less engaging than peer tutoring. This difference in engagement did not account for differences between the peer tutoring and iPad conditions in terms of performance on any of the included measures. This lack of difference in scores between the peer tutoring and technology conditions is consistent with a previous meta-analysis which examined the impact of traditional methods of intervention (e.g., peer-mediated, teacher-mediated, and self-managed) compared to technology-mediated interventions on whole number operations knowledge and found no significant difference between the two intervention types (Kromminga et al., 2020). Within the present study, the lack of identified differences in post-test scores between groups despite the differences in engagement could be due to a lack of statistical power to detect differences or it could be due to similar active ingredients being present across all the intervention conditions.

When examining the list of active ingredients across the iPad and peer tutoring intervention conditions (recognizing that the third, comparison condition, is a combination of the prior two), the only difference is that multiplication facts were typed into the app while the student worked independently on the iPad whereas with peer tutoring students verbalized and wrote their answers. Skinner and colleagues (1997) found that students who verbalized their responses had greater growth in math fluency than those who did not verbalize their answers. This data suggest that the peer tutoring group may have had an advantage in verbalizing their responses, however this did not translate to differences in growth on the included measures. Perhaps other variables discussed above such as the potency of opportunities to practice impacted the growth of the peer tutoring condition. In other words the combined group both may have yielded a higher number of opportunities to respond working on the iPad and were more engaged in the peer-tutoring intervention perhaps producing the small difference between effect sizes on the paper-pencil CBM between the peer-tutoring and combined conditions.

Intervention Acceptability

Each group of students rated their perspective intervention as being highly acceptable. These results are consistent with other studies that have examined student acceptability of technology-delivered interventions (Kromminga & Codding, 2020; Musti Rao & Plati, 2015) and peer tutoring interventions (Greenwood et al., 2001; Maheady et al., 1991; Maheady et al., 2004). There were no significant differences between the conditions in terms of student acceptability, indicating that any of the three intervention conditions might be an acceptable intervention options for students. The classroom teachers also rated that they were very likely to use both the technology-delivered and peer tutoring interventions in their classroom in the future. Previous researchers have found that interventions are generally more acceptable if they are not excessively time consuming (Witt et al., 1984; Witt & Martens, 1983) and have limited or no negative side effects (Kazdin, 1981). Therefore, it is possible that the limited time burden on the teachers for both intervention procedures and the short period of class time required to complete the intervention (e.g., 20 minutes) led to high acceptability among classroom teachers.

Limitations and Future Research

As with all research, the results must be taken with the limitations. First, the sample size for the present study was smaller than the power analysis indicated would be necessary to detect an effect. Future research should ensure that adequate power is achieved to detect differences between groups. Second, the present study did not include a control condition. While this limits the claims we could make regarding the impacts of the intervention compared to students receiving no treatment, researchers have stressed the importance of including active treatment comparisons to provide more information regarding how two treatments compare in terms of instructional design principles, impacts on student performance, and resources for implementation (Gersten et al., 2000; Lewis-Snyder et al., 2002; Petersen-Brown et al., 2019; Pressley & Harris, 1994). Future research may consider including an active control and business as usual control condition to provide comprehensive support for the effectiveness of the treatments. Third, the current sample of 3rd grade students was predominantly white students without disabilities from a middle-class suburb in the Midwest U.S. This limits the generalizability of the results to other grade levels, races or ethnicities, disability status, urbanicity, geographical region, or schools with different instructional practices. Future research should replicate this study with different populations of students to facilitate the generalization of these results to other populations. Fourth, due to limitations with how the facts could be presented in the iPad condition (i.e., all 91 facts randomized) the peer condition also practiced daily with the full set of facts randomized. This was done to keep the intervention conditions similar for comparison between conditions, however, it is often recommended to practice facts in small sets of unknown facts rather than all at once (Burns et al., 2016). Therefore, it is possible that the effects of the interventions could be improved with smaller sets being practiced daily. An additional limitation of the present study was that no interobserver data was collected on the measure of student engagement. Therefore, the reliability of those results is unknown. Additionally, researchers might consider controlling the number of opportunities to respond rather than time engaged in intervention.

Implications for Practice

The results of this study suggest that all students improved their multiplication fluency from pre-test to post-test and there were few differences between groups. Although, the results were non-significant for most comparisons between intervention conditions, a significant difference was yielded for the iPad condition compared to the combined condition on the iPad CBM. Taken together, these results have practical meaning given that reciprocal peer tutoring, iPad delivered flashcards, or the combination of the two were beneficial for improving the multiplication fact fluency of a sample of third grade students from a high achieving school district in the Midwest. Moreover, the interventions were highly acceptable to both the students and teachers. In this study, there was an observed difference in student engagement favoring the peer tutoring intervention, indicating that peer practice may influence student engagement more than technology-mediated practice.

CHAPTER 4

Synthesis and General Discussion

Proficiency with whole number operations is a critical skill for students according to the National Mathematics Advisory Panel (2008). However, results of national and international assessments indicate that student's in the U.S. are not meeting proficiency standards with whole number operations (NCES; 2015, 2019). This is problematic due to the relationship between proficiency with whole numbers and later success in advanced mathematics such as algebra, rational number computation, and word problem solving (Fuchs et al., 2006; Fuchs et al., 2016, Namkunk et al., 2018; NMAP, 2008). Moreover, individuals who have not developed proficiency with whole number operations struggle in adult life. For example, students who fail to master whole number operations are (a) less likely to attend college, (b) make less money, (c) less civically engaged, and (d) more likely to struggle to make economic decisions (Adelman, 2006; Ancker & Kaufman, 2007; Attewell & Domina, 2008; Codding et al., 2017; Crowe, 2010; Dowker, 2005).

Despite having access to years of research which support the use of interventions incorporating frequent opportunities to practice and immediate corrective feedback (Baker et al., 2002; Burns et al., 2006; Codding et al., 2011; Fuchs et al., 2008; Daly et al., 2007; NMAP, 2008); these interventions are not frequently implemented (Spectrum K12, 2011). The lack of implementation of evidence-based interventions is likely due to the numerous implementation barriers faced by schools across the nation such as lack of time, personnel, training, and materials to implement such interventions (Long et al., 2016). This highlights a need to identify effective and feasible interventions that can be implemented at a class wide level to support students who are struggling to master whole number operations.

The two studies in this dissertation aimed to address the need for effective interventions that target student whole number operations knowledge while also alleviating some common barriers to implementation by (a) conducting a meta-analytic review of the literature which has compared two types of intervention modality and (b) examining and comparing the effects of three class wide intervention packages on student whole number operations knowledge. Study 1 synthesized 14 studies comparing a technology-mediated intervention to an active control condition. The instructional principles within each intervention condition and quality of the studies were also evaluated. Study 2 examined the effectiveness of three interventions on student multiplication fluency and compared the effects of these intervention conditions across multiple proximal and distal measures of mathematics achievement. The intervention conditions examined were reciprocal peer tutoring, iPad delivered flashcard drill, and a combined intervention with three days per week of iPad practice and 1 day per week of peer tutoring practice. The findings from these two studies replicated aspects of previous meta-analyses and intervention modality comparison studies.

Study 1 General Findings

Study 1 used meta-analytic techniques to examine studies which have evaluated the impact of technology-mediated intervention on student whole number knowledge compared to a traditional method (i.e., peer-mediated, self-managed, or teacher-mediated). The most common technology used to deliver the technology-mediated interventions across the 14 included studies was computers whereas the most common modality of traditional intervention delivery was a self-managed intervention. The average treatment effect for the technology-mediated interventions from pre- to post-test ($g = 1.22$) was slightly larger than the treatment effect for the traditional modes of intervention delivery ($g = 0.978$). Moreover, three moderators were included in the analyses published, year of publication, and intervention modality. Published studies had higher effect sizes on average than unpublished studies but the difference was not significant (g= -0.333, 95% CI [-1.781, 1.12], p = 0.5923). Studies published between 2000 and 2017 had higher effect sizes on average than those published prior to 2000 (i.e., 1985-1999), however the difference was not significant (g= -0.527, 95% CI [-1.384, 0.33], p = 0.1921). Lastly, technology-mediated interventions had higher effect sizes than traditional modes of intervention delivery, though the difference was not significant (g= 0.118, 95% CI [-0.349, 0.586], p = 0.5734).

While the results of study 1 indicate that either mode of intervention delivery; technology or traditional; may represent viable options for intervention delivery when the goal is to improve whole number basic computation skills. The results of study 1 indicate the mode of intervention delivery may not have significant impacts on student outcomes when accounting for the presence of good instructional design principles. This is consistent with previous researchers who have theorized that technology is merely a means of delivering instructional practices rather than an additional instructional benefit (Clark; 19833, 1985, 1994; Twyman & Sota, 2016). Study 1 was limited by the poor quality of the included studies. Specifically, the included studies lacked information regarding the participant demographics, equivalence between groups at pre-test, and inclusion of procedural fidelity which limit the claims that can be made regarding the equivalence of technology-mediated and traditional methods of intervention. Therefore, while the results of study 1 indicate that there may be no difference between these modes of intervention delivery on student whole number operations knowledge, more high-quality research studies are needed.

Study 2 General Findings

The primary purpose of study 2 was to examine the effectiveness of a peer-mediated, technology-mediated, and combined intervention on the multiplication fact fluency of 3rd grade students.

The results of the present study indicate that the technology-mediated, peer-mediated, and combined conditions showed significant improvement in multiplication fact fluency from pre- to post-test on both proximal measures (e.g., paper-pencil and iPad-based) with moderate to large effect sizes from pre- to post-test for each of the three treatment conditions. On the distal measures of achievement, the treatment conditions had negligible to moderate effect sizes from pre- to post-test. On three of the four assessments there was no significant difference between the treatment conditions at post-test. On the iPad CBM, the iPad condition significantly outperformed the combined condition. In terms of engagement and off-task behavior, students who engaged in reciprocal peer tutoring had significantly higher engagement and lower percentages of off-task behavior than students who engaged in iPad delivered flashcards. Overall, study 2 provided evidence for the use of reciprocal peer tutoring, iPad delivered flashcards, or a combination of the two to improve student multiplication fluency. All groups had significant growth from pre- to post-test on proximal measures which indicates that educators may choose any of these interventions depending on the needs and resources of their school.

Practice and Research Implications

The results of this dissertation have clear implications for practice and future research. First, in study 1 we found that the included studies did not meet quality standards set forth by Jitendra and colleagues (2011). Future research should ensure that the demographics of the participants are reported, treatment fidelity is monitored and reported, and equivalence between pre-test groups is established. Additionally, future analyses should consider examining the impacts of each type of traditional method of intervention separately, study 1 did not have enough power to do so. Therefore, more primary studies that compare the impacts of two different intervention modalities targeting whole number operations are needed.

Study 2 provided evidence to suggest that reciprocal peer tutoring, iPad delivered flashcards, or a combined condition may improve the multiplication fluency of predominantly white third grade students from a high achieving suburban school district in the Midwestern U.S. Future research should replicate this study with different grade levels, races or ethnicities, disability status, urbanicity, geographical region, or schools with different instructional practices. Additionally, future research may wish to consider including a business as usual control condition to provide support for the effectiveness of the treatments compared to a control condition. Future research should also consider measuring opportunities to respond and might consider controlling the number of opportunities to respond rather than time spent in intervention.

Practically, the results of this dissertation indicate that either technology-mediated or traditional methods of intervention may be effective at improving the whole number operations knowledge of 2nd to 5th grade students in the U.S. Moreover, practitioners should carefully examine potential intervention practices for the presence of instructional design principles prior to their implementation with students. Lastly, because either mode of intervention delivery may be effective educators should carefully examine the resources which are available within their school and first implement interventions which utilize available resources.

Conclusion

This dissertation aimed to address the need for effective interventions which target whole number operations knowledge while also minimizing common implementation barriers. Study 1 analyzed literature which compared a technology-mediated intervention to a traditional method of intervention on student whole number operations knowledge. While there were multiple strengths and limitations of the meta-analysis, the main finding suggests that there were no significant differences between technology-mediated interventions and traditional methods of intervention on student number operations knowledge. This suggests that educators may choose either method to improve student whole number operations knowledge but should also consider the presence of common instructional design principles. Study 2 provided additional support for the use of either a technology-mediated intervention, peer-mediated intervention, or a combined intervention to improve multiplication fact fluency. Results indicate that all three treatment conditions had significant growth from pre- to post-test on proximal paper-pencil and iPad-based measures of fact fluency. However, students in the iPad condition outperformed students in the combined condition on the iPad CBM. Knowledge regarding the equivalence of effects on student math performance and effective instructional design principles across intervention modalities will help educators make decisions regarding the allocation of resources to improve the whole number operations knowledge of their students while accounting for the resources available to them to implement the intervention, thereby alleviating common implementation barriers.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Adelman, C. (2006). The toolbox revisited: Paths to degree completion from high school through college. US Department of Education.

Ancker, J. S., & Kaufman, D. (2007). Rethinking health numeracy: A multidisciplinary literature review. Journal of the American Medical Informatics Association, 14, 713-721.

Attewell, P., & Domina, T. (2008). Raising the bar: Curricular intensity and academic performance. Educational Evaluation and Policy Analysis, 30, 51-71.

Averill, O. H., Rinaldi, C., & Collaborative, U. S. E. L. (2011). Multi-tier system of supports (MTSS). District Administration, 48(8), 91-95.

Baker, S., Gersten, R., & Lee, D. S. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. The Elementary School Journal, 103(1), 51-73.

Begeny, J. C. (2009). Helping early literacy with practice strategies (HELPS): A one-on-one program designed to improve students' reading fluency. Raleigh, NC: The Helps Education Fund. Retrieved from http://www.helpsprogram.org.

Berch, D. B., & Mazzocco, M. M. (2007). Why is math so hard for some children. The nature and origins of mathematical learning difficulties and disabilities. Baltimore, MD: Brookes.

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. The Behavior Analyst, 19(2), 163-197.

Bingimlas, K. A. (2009). Barriers to the successful integration of ICT in teaching and learning environments: A review of the literature. *Eurasia Journal of Mathematics, science and technology education, 5*(3), 235-245.

Bliss, S. L., Skinner, C. H., Hautau, B., & Carroll, E. E. (2008). Articles published in four school psychology journals from 2000 to 2005: An analysis of experimental/intervention research. *Psychology in the Schools, 45*(6), 483-498.

Borenstein, M. (2009). *Effect sizes for continuous data*. In Cooper, H., Hedges, L., & Valentine, J. (Eds.) The handbook of research synthesis and meta-analysis, 221-235.

Bowman-Perrott, L., Davis, H., Vannest, K., Williams, L., Greenwood, C., & Parker, R. (2013). Academic benefits of peer tutoring: A meta-analytic review of single-case research. *School psychology review, 42*(1), 39-55.

Bryant, B. R., Bryant, D. P., Kethley, C., Kim, S. A., Pool, C., & Seo, Y. J. (2008). Preventing mathematics difficulties in the primary grades: The critical features of instruction in textbooks as part of the equation. *Learning Disability Quarterly, 31*(1), 21-35.

Bryant, B. R., Ok, M., Kang, E. Y., Kim, M. K., Lang, R., Bryant, D. P., & Pfannestiel, K. (2015). Performance of fourth-grade students with learning disabilities on multiplication facts comparing teacher-mediated and technology-mediated interventions: A preliminary investigation. *Journal of Behavioral Education, 24,* 255-272.

Burke, Robert L., (1982). *CAI Sourcebook.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Burns, M. K., Kanive, R., & DeGrande, M. (2012). Effect of a computer-delivered math fact intervention as a supplemental intervention for math in third and fourth grades. *Remedial and Special Education, 33*(3), 184-191.

Burns, M. K., VanDerHeyden, A. M., & Jiban, C. L. (2006). Assessing the instructional level for mathematics: A comparison of methods. *School Psychology Review, 35*(3), 401.

Burns, M. K., Ysseldyke, J., Nelson, P. M., & Kanive, R. (2015). Number of repetitions required to retain single-digit multiplication math facts for elementary students. *School Psychology Quarterly, 30*, 398.

Burns, M. K., Zaslofsky, A. F., Maki, K. E., & Kwong, E. (2016). Effect of modifying intervention set size with acquisition rate data while practicing single-digit multiplication facts. *Assessment for Effective Intervention, 41*(3), 131-140.

Burns, P. K., & Bozeman, W. C. (1981). Computer-assisted instruction and mathematics achievement: Is there a relationship?. *Educational Technology, 21*(10), 32-39.

*Campbell, D. L., Peck, D. L., Horn, C. J., & Leigh, R. K. (1987). Comparison of computer-assisted instruction and print drill performance: A research note. *ECTJ, 35,* 95-103.

Carr, M., & Alexeev, N. (2011). Fluency, accuracy, and gender predict developmental trajectories of arithmetic strategies. *Journal of Educational Psychology, 103*(3), 617.

Carr, M., Steiner, H. H., Kyser, B., & Biddlecomb, B. (2008). A comparison of predictors of early emerging gender differences in mathematics competency. *Learning and Individual Differences, 18*(1), 61-75.

*Carrier, C., Post, T. R., & Heck, W. (1985). Using microcomputers with fourth-grade students to reinforce arithmetic skills. *Journal for Research in Mathematics Education, 16*(1), 45-51.

Cates, G. L. (2005). Effects of peer versus computer assisted drill on mathematics response rates. *Psychology in the Schools, 42,* 637-646.

Chafouleas, S. M., & Riley Tillman, T. C. (2005). Accepting the gap: An introduction to the special issue on bridging research and practice. *Psychology in the Schools, 42*(5), 455-458.

Cheung, A. C., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational research review, 9*, 88-113.

Christ, T. J., Scullin, S., Tolbize, A., & Jiban, C. L. (2008). Implications of recent research: Curriculum-based measurement of math computation. *Assessment for Effective Intervention, 33*(4), 198-205.

Clark, R. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53,* 445–459.

Clark, R. E. (1985). Evidence for confounding in computer-based instruction studies: Analyzing the meta-analyses. *Educational Communications and Technology Journal, 33*(4), 249–262.

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development, 42*(2), 21–29.

Clark, R. E., & Salomon, G. (1986). *Media in teaching*. In M. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 464–478). New York: Macmillan.

Clarke, B., Doabler, C. T., Smolkowski, K., Turtura, J., Kurtz-Nelson, E., Fien, H., & Baker, S. (2019). Exploring the relationship between initial math skill and a kindergarten mathematics intervention. *Exceptional Children, 85,*129–146.

Codding, R. S., & Poncy, B. C. (2010). Introduction to the special issue: Toward an explicit technology for generalizing academic behavior. *Journal of Behavioral Education, 19*(1), 1-6.

Codding, R. S., Burns, M. K., & Lukito, G. (2011). Meta□analysis of mathematic basic□ fact fluency interventions: A component analysis. *Learning Disabilities Research & Practice, 26*, 36-47.

Codding, R. S., Volpe, R. J., & Poncy, B. C. (2017). *Effective math interventions: A guide to improving whole-number knowledge*. Guilford Publications.

Cohen, J. (1962). The statistical power of abnormal–social psychological research: A review. *Journal of Abnormal and Social Psychology, 65*, 145–153.

Cohen, P. A., Kulik, J. A., & Kulik, C. L. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19*(2), 237-248.

Collier-Meek, M. A., Fallon, L. M., Sanetti, L. M., & Maggin, D. M. (2013). Focus on implementation: Assessing and promoting treatment fidelity. *Teaching Exceptional Children, 45(*5), 52-59.

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). Russell Sage Foundation.

Crowe, A. R. (2010). “What's Math Got to Do With It?”: Numeracy and Social Studies Education. *The Social Studies, 101*, 105-110.

Daly, E. J., Martens, B. K., Barnett, D., Witt, J. C., & Olson, S. C. (2007). Varying intervention delivery in response to intervention: Confronting and resolving challenges with measurement, instruction, and intensity. *School Psychology Review, 36*, 562.

Dineen, J. P., Clark, H. B., & Risley, T. R. (1977). Peer tutoring among elementary students: educational benefits to the tutor 1. *Journal of applied behavior analysis, 10*(2), 231-238.

Doabler, C. T., Cary, M. S., Jungjohann, K., Clarke, B., Fien, H., Baker, S., ... & Chard, D. (2012). Enhancing core mathematics instruction for students at risk for mathematics disabilities. *Teaching Exceptional Children, 44*, 48-57.

Dowker, A. (2005). Early identification and intervention for students with mathematics difficulties. *Journal of learning disabilities, 38*, 324-332.

Dufrene, B. A., Noell, G. H., Gilbertson, D. N., & Duhon, G. J. (2005). Monitoring implementation of reciprocal peer tutoring: Identifying and intervening with students who do not maintain accurate implementation. *School Psychology Review, 34*(1), 74-86.

*Duhon, G. J., House, S. H., & Stinnett, T. A. (2012). Evaluating the generalization of math fact fluency gains across paper and computer performance modalities. *Journal of school psychology, 50,* 335-345.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... & Sexton, H. (2007). School readiness and later achievement. *Developmental psychology, 43*, 1428.

Eckert, T. L., Hier, B. O., Hamsho, N. F., & Malandrino, R. D. (2017). Assessing children's perceptions of academic interventions: The Kids Intervention Profile. *School Psychology Quarterly, 32*(2), 268.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*(7109), pp.629-634.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior research methods, instruments, & computers, 28*(1), 1-11.

Ertmer, P. (1999). Addressing first- and second-order barriers to change: Strategies for technology implementation. *Educational Technology Research and Development, 47*(4), 47–61.

Evans, M. A., Nino, M., Deater-Deckard, K., & Chang, M. (2015). School-wide adoption of a mathematics learning game in a middle school setting: Using the TPACK framework to analyze effects on practice. *The Asia Pacific Education Researcher, 24,* 495–504.

Fantuzzo, J. W., King, J. A., & Heller, L. R. (1992). Effects of reciprocal peer tutoring on mathematics and school adjustment: A component analysis. *Journal of Educational Psychology, 84*(3), 331.

Fantuzzo, J. W., Riggio, R. E., Connelly, S., & Dimeff, L. A. (1989). Effects of reciprocal peer tutoring on academic achievement and psychological adjustment: A component analysis. *Journal of educational psychology*, *81*(2), 173.

Fantuzzo, J., & Ginsburg-Block, M. (1998). Reciprocal peer tutoring: Developing and testing effective peer collaborations for elementary school students. *Peer-assisted learning*, 121-144.

*Fletcher-Finn, C., & Gravatt, B. (1995). The efficacy of computer-assisted instruction (CAI): A meta-analysis. *Journal of Educational Computing Research, 12*, 219–241.

*Flynn, L. (2010). Computer assisted instruction or mnemonic strategy instruction: which approach produces better outcomes with helping students develop multiplication fact fluency?. (Doctoral dissertation) Retrieved from ProQuest Dissertations Publishing, 2011.

Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring measures in mathematics: A review of the literature. *The Journal of Special Education, 41*(2), 121-139.

FormSoft Group LTD (2012). Timed Test [Mobile application software]. Retrieved from https://apps.apple.com/us/app/timed-test/id497960697

Franca, V. M., Kerr, M. M., Reitz, A. L., & Lambert, D. (1990). Peer tutoring among behaviorally disordered students: Academic and social benefits to tutor and tutee. *Education and Treatment of Children*, 109-128.

Fuchs, D., & Fuchs, L.S. (2017). Critique of the National Evaluation of Responsiveness-To-Intervention: A case for simpler frameworks. *Exceptional Children, 83*, 255–268.

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., ... & Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98*, 29.

Fuchs, L. S., Fuchs, D., Powell, S. R., Seethaler, P. M., Cirino, P. T., & Fletcher, J. M. (2008). Intensive intervention for students with mathematics disabilities: Seven principles of effective practice. *Learning Disability Quarterly, 31,* 79–92.

Fuchs, L. S., Gilbert, J. K., Powell, S. R., Cirino, P. T., Fuchs, D., Hamlett, C. L., ... & Tolar, T. D. (2016). The role of cognitive processes, foundational math skill, and calculation accuracy and fluency in word-problem solving versus prealgebraic knowledge. *Developmental Psychology, 52*, 2085.

*Fuson, K. C., & Brinko, K. T. (1985). The comparative effectiveness of microcomputers and flash cards in the drill and practice of basic mathematics facts. *Journal for Research in Mathematics Education*, 225-232.

Gersten, R., & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *The Journal of Special Education*, *33*(1), 18-28.

Gersten, R., Baker, S., & Lloyd, J. W. (2000). Designing high-quality research in special education: Group experimental design. The Journal of Special Education, 34(1), 2-18.

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009). Assisting Students Struggling with Mathematics: Response to Intervention (RtI) for Elementary and Middle Schools. NCEE 2009-4060. *What Works Clearinghouse*.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional children, 71*, 149-164.

*Gesbocker, S. R. (2011). *The effects of computer assisted instruction on the recall of basic math facts among middle school students.* (Doctoral dissertation) Retrieved from ProQuest Dissertations Publishing, 2011.

Ginsburg-Block, M. D., Rohrbeck, C. A., & Fantuzzo, J. W. (2006). A meta-analytic review of social, self-concept, and behavioral outcomes of peer-assisted learning. *Journal of Educational Psychology, 98*, 732-749.

*Gmitter, J. W. (1989). *Effects of microcomputer-assisted instruction and class wide peer tutoring on computational skill achievement of third-grade students. (Doctoral dissertation)* Retrieved from ACM Digital Library.

Greenwood, C. R., (1991). Longitudinal analysis of time, engagement, and achievement of at risk versus non risk students. *Exceptional Children, 57*(6),521-535.

Greenwood, C. R., Arreaga-Mayer, C., Utley, C. A., Gavin, K. M., & Terry, B. J. (2001). Class wide peer tutoring learning management system: Applications with elementary-level English language learners. *Remedial and Special Education, 22*(1), 34-47.

Greenwood, C. R., Terry, B., Utley, C.A., Montagna, D., & Walker, IX (1993). Achievement, placement, and services: Middle school benefits of class wide peer tutoring used at the elementary school. *School Psychologist, 22*(30), 497-516

Gustafson, K. L., & Tillman, M. H. (1991). Designing the general strategies of instruction. *Instructional design: Principles and applications*, 173-192.

Haring, N.G., & Eaton, M.D. (1978). *Systematic instructional procedures: An instructional hierarchy.* In Haring, N.G., Lovitt, T.C., Eaton, M.D., & Hansen, C.L. (Eds.), The fourth R: Research in the classroom (pp. 23-40). Columbus OH: Charles E. Merrill Publishing Co.

Harris, K. R. (1988). Learning disabilities research: The need, the integrity, and the challenge. *Journal of learning disabilities, 21*, 267-270.

Hattie, J. A. C., & Learning, V. (2009). *A synthesis of over 800 meta-analyses relating to achievement.* New York.

Haydon, T., Hawkins, R., Denune, H., Kimener, L., McCoy, D., & Basham, J. (2012). A comparison of iPads and worksheets on math skills of high school students with emotional disturbance. *Behavioral Disorders, 37,* 232-243.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107-128.

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods, 3*(3), 224-239.

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2013). A standardized mean difference effect size for multiple baseline designs across individuals. *Research Synthesis Methods, 4*(4), 324-341.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165-179.

Jitendra, A. K., Burgess, C., & Gajria, M. (2011). Cognitive strategy instruction for improving expository text comprehension of students with learning disabilities: The quality of evidence. *Exceptional Children, 77*, 135-159.

Jordan, N. C., Glutting, J., & Ramineni, C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learning and Individual Differences, 20*(2), 82-88.

Jordan, N. C., Resnick, I., Rodrigues, J., Hansen, N., & Dyson, N. (2017). Delaware longitudinal study of fraction learning: Implications for helping children with mathematics difficulties. *Journal of Learning Disabilities, 50*, 621-630.

*Kanive, R., Nelson, P. M., Burns, M. K., & Ysseldyke, J. (2014). Comparison of the effects of computer-based practice and conceptual understanding interventions on mathematics fact retention and generalization. *The Journal of Educational Research, 107*(2), 83-89.

Kazdin, A. E. (1981). Acceptability of child treatment techniques: The influence of treatment efficacy and adverse side effects. *Behavior Therapy, 12*(4), 493-506.

Ketterlin-Geller, L. R., & Chard, D. J. (2011). Algebra readiness for students with learning difficulties in grades 4–8: Support through the study of number. *Australian Journal of Learning Difficulties, 16*, 65-78.

Kiger, D., Herro, D., & Prunty, D. (2012). Examining the influence of a mobile learning intervention on third grade math achievement. *Journal of Research on Technology in Education, 45*(1), 61–82.

Kilgus, S. P., Riley-Tillman, T. C., & Kratochwill, T. R. (2016). Establishing interventions via a theory-driven single case design research cycle. *School Psychology Review, 45*, 477-498.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). The strands of mathematical proficiency. *Adding it up: Helping Children Learn Mathematics,* 115-118.

Kiru, E. W., Doabler, C. T., Sorrells, A. M., & Cooc, N. A. (2018). A synthesis of technology-mediated mathematics interventions for students with or at risk for mathematics learning disabilities. *Journal of Special Education Technology, 33*, 111-123.

Kroesbergen, E. H., & Van Luit, J. E. (2003). Mathematics interventions for children with special educational needs: A meta-analysis. *Remedial and Special Education, 24*, 97-114.

Kromminga, K. R., & Codding, R. S. (2020). A Comparison of 1: 1 Flashcards and a Tablet App on Student Mathematics Proficiency. *Journal of Behavioral Education*

Kromminga, K. R., Codding, R. S., Rios, J., Goodridge, A., & Kember, J. (2020). *Examining Effects of Instructional Delivery and Design Principles on Whole Number Operations: A Meta-Analytic Review.* (Doctoral dissertation).

Kulik, J. A., & Kulik, C. L. C. (1987). Review of recent research literature on computer-based instruction. *Contemporary Educational Psychology, 12*(3), 222-230.

Kunsch, C. A., Jitendra, A. K., & Sood, S. (2007). The effects of peer□mediated instruction in mathematics for students with learning problems: A research synthesis. *Learning Disabilities Research & Practice*, *22*(1), 1-12.

Landeen, J. J., & Adams, D. A. (1988). Computer-assisted drill and practice for behaviorally handicapped learners: proceed with caution. *Education and Treatment of Children,* 218-229.

Lewis-Snyder, G., Stoiber, K. C., & Kratochwill, T. R. (2002). Evidence-based interventions in school psychology: An illustration of task force coding criteria using group-based research design. *School Psychology Quarterly, 17*(4), 423–465.

Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review, 22*, 215-243.

Liao, Y. K. (1998). Effects of hypermedia versus traditional instruction on students' achievement: A meta-analysis. *Journal of Research on Computing in Education, 30*, 341–359.

Long, A. C., Sanetti, L. M. H., Collier-Meek, M. A., Gallucci, J., Altschaefl, M., & Kratochwill, T. R. (2016). An exploratory investigation of teachers' intervention planning and perceived implementation barriers. *Journal of School Psychology, 55*, 1-26.

MacArthur, C. A., & Graham, S. (1987). Learning disabled students' composing under three methods of text production: Handwriting, word processing, and dictation. *The Journal of Special Education, 21*(3), 22-42.

Maheady, L., & Harper, G. F. (1987). A class wide peer tutoring program to improve the spelling test performance of low-income, third-and fourth-grade students. *Education and Treatment of Children*, 120-133.

Maheady, L., Harper, G. F., & Mallette, B. (2001). Peer-mediated instruction and interventions and students with mild disabilities. *Remedial and Special Education*, *22*(1), 4-14.

Maheady, L., Harper, G. F., Mallette, B., & Karnes, M. (2004). Preparing preservice teachers to implement class wide peer tutoring. *Teacher Education and Special Education, 27*(4), 408-418.

Maheady, L., Harper, G. F., Mallette, B., & Winstanley, N. (1991). Training and implementation requirements associated with the use of a class wide peer tutoring system. *Education and Treatment of Children,* 177-198.

*Mattoon, C., Bates, A., Shifflet, R., Latham, N., & Ennis, S. (2015). Examining Computational Skills in Prekindergarteners: The Effects of Traditional and Digital Manipulatives in a Prekindergarten Classroom. Early Childhood Research & Practice, 17(1), n1.

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Technical Manual. Woodcock Johnson IV.* Rolling Meadows, IL: Riverside

McKenna, C. (2012). There's an app for that: How two elementary classrooms used iPads to enhance student learning. *Education, 2,* 136–142.

McNamara, N. (2017). *Mental Math Cards Challenge (Version 1.7.2).* [Mobile application software]. Retrieved from https://apps.apple.com/us/app/mental-math-cards-challenge/id500595081

Meaney, K. (2017). *Apple's iPad is Most Popular Tablet in Schools.* Simba Information. Retrieved from https://www.simbainformation.com/Content/Blog/2017/08/07/Apples-iPad-is-Most-Popular-Tablet-in-Schools

Mellard, D., McKnight, M., & Jordan, J. (2010). RTI tier structures and instructional intensity. *Learning Disabilities Research & Practice, 25*(4), 217-225.

Methe, S. A., Briesch, A. M., & Hulac, D. (2015). Evaluating procedures for reducing measurement error in math curriculum-based measurement probes. *Assessment for Effective Intervention, 40*(2), 99-113.

Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities, 42*, 306-321.

Moyer-Packenham, P. S., & Westenskow, A. (2013). Effects of virtual manipulatives on student achievement and mathematics learning. *International Journal of Virtual and Personal Learning Environments, 4*(3), 35-50.

Musti-Rao, S., & Plati, E. (2015). Comparing two class wide interventions: Implications of using technology for increasing multiplication fact fluency. *Journal of Behavioral Education, 24,* 418-437.

Najarian, K., & Vasilache, S. (2012). PAC Learning. Encyclopedia of the Sciences of Learning, 2545-2546.

Namkung, J. M., Fuchs, L. S., & Koziol, N. (2018). Does initial learning about the meaning of fractions present similar challenges for students with and without adequate whole-number skill?. *Learning and individual differences, 61*, 151-157.

National Center for Education Statistics, National Assessment of Educational Progress (NAEP), *2019 Mathematics Assessment.* Retrieved from https://www.nationsreportcard.gov/ndecore/

National Center for Education Statistics. (2015). *Trends in international mathematics and science study (TIMSS). Institute of Education Sciences.* Retrieved from https://nces.ed.gov/timss/

National Council of Teachers of Mathematics. (2013). *Principles and standards for school mathematics.* Reston, VA: Author

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel.* US Department of Education.

O'Malley, P., Lewis, M. E. B., Donehower, C., & Stone, D. (2014). Effectiveness of using iPads to increase academic task completion by students with autism. *Universal Journal of Educational Research, 2*, 90–97.

Okey, J. R. (1991). Procedures of lesson design. Instructional design: principles and applications. 2nd edition. Englewood Cliffs, New Jersey: *Education Technology Publications,* 192-208.

Okolo, C. M. (1992). The effect of computer□assisted instruction format and initial attitude on the arithmetic facts proficiency and continuing motivation of students with learning disabilities. Exceptionality: *A Special Education Journal, 3*(4), 195-211.

Patton, J. R., Cronin, M. E., Bassett, D. S., & Koppel, A. E. (1997). A life skills approach to mathematics instruction: Preparing students with learning disabilities for the real-life math demands of adulthood. *Journal of Learning Disabilities, 30*(2), 178-187.

Pearson, N. C. S. (2011). *AIMSweb default cut scores explained*. Bloomington, MN: Author.

Petersen□Brown, S. M., Henze, E. E., Klingbeil, D. A., Reynolds, J. L., Weber, R. C., & Codding, R. S. (2019). The use of touch devices for enhancing academic achievement: A meta□analysis. *Psychology in the Schools, 56*(7), pp.1187-1206.

*Pflug, E. A. (1987). *A comparison of the effect of computer-assisted-instruction and same-age peer-tutoring on math achievement of fourth grade students.* (Doctoral

Dissertation). Retrieved from https://ir.library.oregonstate.edu/concern/graduate_thesis_or_dissertations/

*Podell, D. M., Tournaki-Rein, N., & Lin, A. (1992). Automatization of mathematics skills via computer-assisted instruction among students with mild mental handicaps. Education and Training in Mental Retardation, 200-206.

Poncy, B. C., Skinner, C. H., & Jaspers, K. E. (2007). Evaluating and comparing interventions designed to enhance math fact accuracy and fluency: Cover copy, and compare versus taped problems. Journal of Behavioral Education, 16, 27–37

Pressley, M., & Harris, K. R. (1994). Increasing the quality of educational intervention research. Educational Psychology Review, 6, 191-208.

Price, G. R., Mazzocco, M. M., & Ansari, D. (2013). Why mental arithmetic counts: brain activation during single digit arithmetic predicts high school math scores. Journal of Neuroscience, 33(1), 156-163.

R Core, T. E. A. M. (2016). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rhymer, K. N., Dittmer, K. I., Skinner, C. H., & Jackson, B. (2000). Effectiveness of a multi-component treatment for improving mathematics fluency. School Psychology Quarterly, 15(1), 40.

*Rich, S. E. H., Duhon, G. J., & Reynolds, J. (2017). Improving the Generalization of Computer-Based Math Fluency Building Through the Use of Sufficient Stimulus Exemplars. Journal of Behavioral Education, 26, 123-136.

Robinson, D. R., Schofield, J. W., & Steers-Wentzell, K. L. (2005). Peer and cross-age tutoring in math: Outcomes and their design implications. *Educational Psychology Review, 17*(4), 327-362.

Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology*, *95*(2), 240.

Rosen, Y., & Salomon, G. (2007). The differential learning achievements of constructivist technology-intensive learning environments as compared with traditional ones: A meta-analysis. *Journal of Educational Computing Research, 36*(1), 1-14.

Sanetti, L. M. H., Gritter, K. L., & Dobey, L. (2011). Treatment integrity of interventions with children in the school psychology literature from 1995 to 2008. *School Psychology Review, 40,* 72–84.

Schrank, F. A., McGrew, K. S., Mather, N., Wendling, B. J., & LaForte, E. M. (2014). *Woodcock-Johnson IV Tests of achievement: Woodcock-Johnson IV tests of cognitive abilities.* Riverside Publishing Company.

Seo, Y. J., & Bryant, D. P. (2009). Analysis of studies of the effects of computer-assisted instruction on the mathematics performance of students with learning disabilities. *Computers & Education, 53*(3), 913-928.

Shapiro, E. S. (1996). *Academic skills problems workbook.* New York: Guilford Press.

Shapiro, E. S. (2011). *Academic skills problems: Direct assessment and intervention*. Guilford Press.

Shinn, M. R. (2004). *Administration and Scoring Manual CBM.* AIMSweb.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis.* New York: Appleton-Century.

Skinner, C. H., & Daly, E. J. (2010). Improving generalization of academic skills: Commentary on the special issue. *Journal of Behavioral Education, 19*(1), 106-115.

Skinner, C. H., Belfiore, P. B., & Watson, T. S. (1995/2002). Assessing the relative effects of interventions in students with mild disabilities: Assessing instructional time. *Journal of Psychoeducational Assessment, 20,* 345–356 (Reprinted from *Assessment in Rehabilitation and Exceptionality, 2*, 207–220.)

Skinner, C. H., Belfiore, P. J., Mace, H. W., Williams-Wilson, S., & Johns, G. A. (1997). Altering response topography to increase response efficiency and learning rates. *School Psychology Quarterly, 12*(1), 54.

Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research, 78*(3), 427-515.

Slavin, R. E., Lake, C., & Groff, C. (2009). Effective programs in middle and high school mathematics: A best-evidence synthesis. *Review of Educational Research, 79*, 839-911.

*Smith, C. L. (2010). *Examining the effectiveness of peer-tutoring and computer-aided instruction for mastery of multiplication facts.* (Doctoral dissertation). Retrieved from https://www.learntechlib.org/p/124963/.

Smith, M., Glass, G., & Miller, T. (1980). *The benefits of psychotherapy*. Baltimore: Johns Hopkins University Press.

Sokolowski, A., Li, Y., & Willson, V. (2015). The effects of using exploratory computerized environments in grades 1 to 8 mathematics: A meta-analysis of research. *International Journal of STEM Education, 2*(1), 8.

Spectrum K12 School Solutions. (2011). *Response to Intervention Adoption Survey 2011.* Retrieved from http://sss.usf.edu/resources/presentations/2010/fasp_summer_inst2010/Resource_SLD/RTI/2010RTIAdoptionSurveyReport.pdf

Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology, 105*(4), 970.

Stenhoff, D. M., & Lignugaris/Kraft, B. (2007). A review of the effects of peer tutoring on students with mild disabilities in secondary settings. *Exceptional children, 74*(1), 8-30.

Stickney, E. M., Sharp, L. B., & Kenyon, A. S. (2012). Technology-enhanced assessment of math fact automaticity: Patterns of performance for low-and typically achieving students. *Assessment for Effective Intervention, 37,* 84-94.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization 1. *Journal of Applied Behavior Analysis, 10(*2), 349-367.

Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study. *Review of Educational research, 81*, 4-28.

Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics, 27*(1), 77-83.

Tipton, E. (2014). How generalizable is your experiment? An index for comparing experimental samples and populations. *Journal of Educational and Behavioral Statistics, 39*(6), 478-501.

Twyman, J. S., & Sota, M. S. (2016). *Educational technology and response to intervention: Affordances and considerations.* In Handbook of Response to Intervention (pp. 493-517). Springer, Boston, MA.

VanDerHeyden, A., McLaughlin, T., Algina, J., & Snyder, P. (2012). Randomized evaluation of a supplemental grade-wide mathematics intervention. *American Educational Research Journal, 49*(6), 1251-1284.

Williams, V. S., Jones, L. V., & Tukey, J. W. (1999). Controlling error in multiple comparisons, with examples from state-to-state differences in educational achievement. *Journal of Educational and Behavioral Statistics, 24*(1), 42-69.

Witt, J. C., & Martens, B. K. (1983). Assessing the acceptability of behavioral interventions used in classrooms. *Psychology in the Schools, 20*(4), 510-517.

Witt, J. C., Elliott, S. N., & Martens, B. K. (1984). Acceptability of behavioral interventions used in classrooms: The influence of amount of teacher time, severity of behavior problem, and type of intervention. *Behavioral Disorders, 9*(2), 95-104.

Young, J. (2017). Technology-enhanced mathematics instruction: A second-order meta-analysis of 30 years of research. *Educational Research Review, 22*, 19-33.

Table 1
Participant and Setting Features
Participants Demographics

Table 2
Primary Outcome Measure Features

Assessment Features

Table 3
Intervention Features by Mode of Intervention Delivery

Technology Intervention Features

Table 4
Instructional design principles

Table 5
Quality Indicators

Study

Table 6
Effect Sizes

Study

Table 7
Core Intervention Components Between iPad and Peer Tutoring Conditions

Comparing Intervention Components

Table 8
Within Condition Statistics Across Measures

Figure 1
Prisma Chart of Systematic Inclusion for Study 1

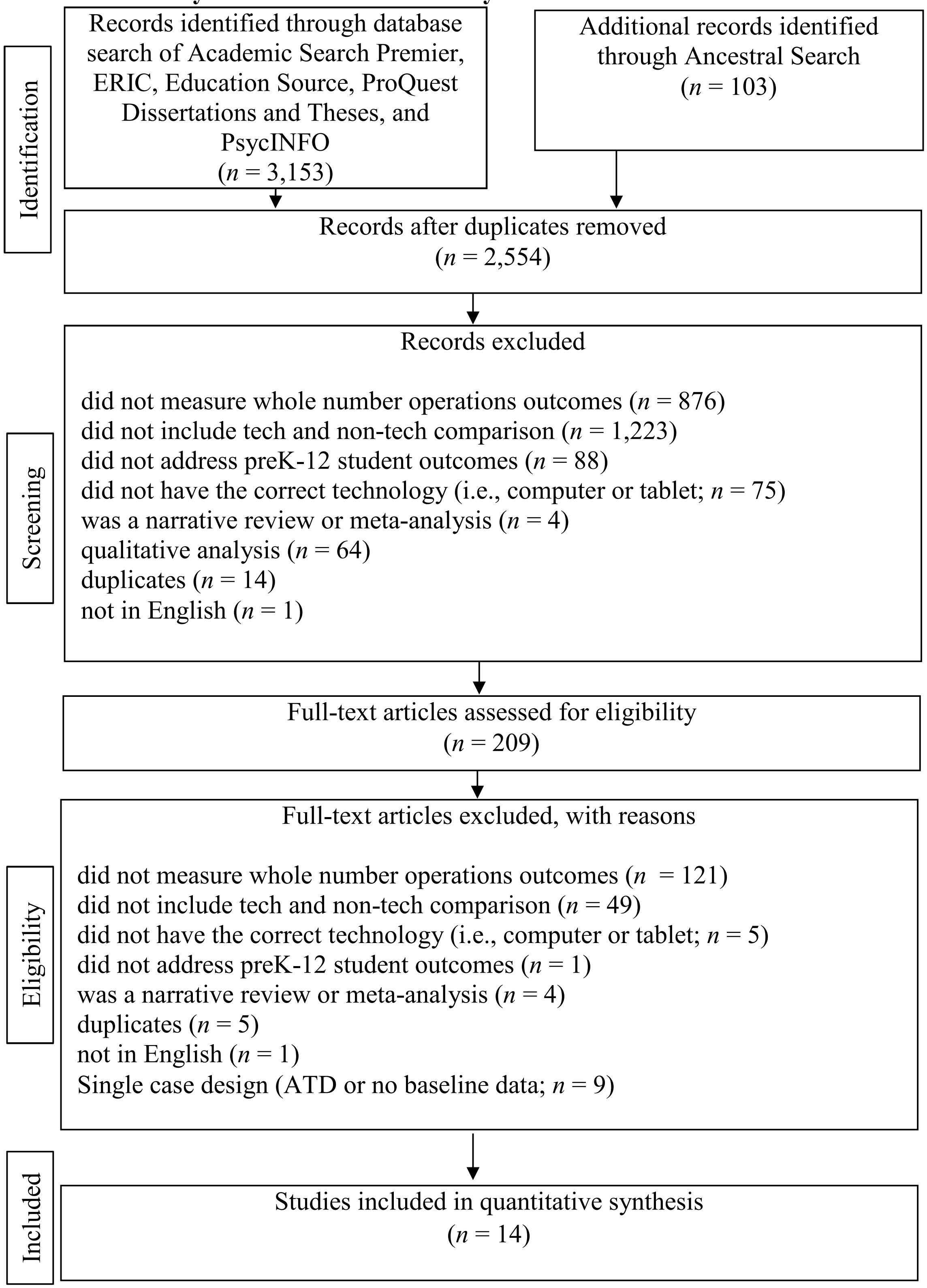
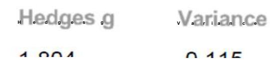


Figure 2

Funnel Plot of Mean Differences and Inverse Variance of Effect Sizes Included in Study 1

Figure 3

Study 1 Forest Plot of Hedges g Effect Sizes and Variances by Study and Intervention Modality



Appendices

Appendix A

Summary of Search Terms and Specifiers Used Across Databases Study 1

Dataset

Appendix B

Screenshot of Timed Test App Interface

Appendix C

What-I-Know Mat used by Students Practicing through Peer Tutoring

Appendix D

Screenshot of Mental Math Cards Challenge App Interface

Appendix E

Screenshot of Practice Statistics Presented to Student After Completing an iPad Flashcard Round