The Effects of Computer-Assisted Instruction in Reading: A Meta-Analysis

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Amy K. Kunkel

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

Dr. Kristen McMaster, Adviser

May, 2015

Acknowledgements

When I began my graduate studies in Special Education in the Educational Psychology Department at the University of Minnesota, I was told by my adviser, Dr. Kristen McMaster that it takes a village to succeed in the doctoral program. Without my village, none of this would have been possible. I would like to express my deepest gratitude to Dr. McMaster, who served as my guide and mentor through the completion of this project and throughout my graduate school experience. Her wealth of knowledge and willingness to help me develop my understanding of research and writing have propelled my learning further than I ever thought possible.

I am also extremely grateful to Dr. Jennifer McComas, who modeled effective leadership in community partnerships and lent her time to help my colleagues and me develop skills in experimental research. Dr. McComas served as the principal investigator of the Leadership Training Grant, which funded my work in this program and allowed me to have valuable experiences in educational leadership.

I would like to thank Dr. Michael Harwell, who inspired this study and provided his expertise and guidance throughout the project. Dr. Matthew Burns, whose passion for best practice in education is infectious, was also instrumental in my completion of this project and the program as a whole. The knowledge and insights imparted by my committee members and professors is immeasurable, and I am deeply grateful for their attentiveness and assistance in the development of this dissertation.

Finally, I would like to thank Dr. Ann Ryan, who encouraged me to pursue my dreams and graciously offered support, guidance, and friendship along the way.

Dedication

I dedicate my dissertation to my family, the core of my village, especially…

…to my husband, Kris, for your unwavering support and selfless sacrifice

…to Isaac, Liam, and Claire, for your unconditional love and hilarious antics

…to my parents and siblings for making me who I am today

Abstract

The purpose of this study was to investigate the effectiveness of computer-assisted instruction (CAI) to improve the reading outcomes of students in preschool through high school. A total of 61 studies met criteria for this review, and 101 independent effect sizes were extracted. Results indicated that the mean effects for students receiving reading CAI were small, positive, and statistically significant when compared to control groups receiving no treatment or non-reading CAI. Categorical moderator analyses and meta-regression were conducted to explore the variation in effects. Results of an analysis of research quality indicated that, on average, about half of quality indicators were met. The results of this meta-analysis show that CAI in reading can effectively enhance the reading outcomes of students in preschool through high school. Future, high-quality research should be conducted to identify effective programs and establish best practice in the instructional design of CAI to enhance the reading skills of all students.

Table of Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

Challenges in Reading

A recent report from the National Assessment of Educational Progress (NAEP; National Council of Educational Statistics [NCES], 2014) indicates a need for targeted interventions in reading for struggling students, with 32% of fourth-grade students and 22% of eighth-grade students scoring below the basic level of proficiency. Though trends on the NAEP point to large gains in reading proficiency over the last 20 years, only 35% of fourth-grade students and 36% of eighth-grade students were considered proficient in reading on the 2013 administration. Of those students not proficient in reading by third grade, three quarters will remain poor readers in high school and most of these students will eventually drop out before earning their diploma. Without early intervention, poor readers face negative outcomes in adulthood, including lower earnings, unemployment, and poverty (Annie E. Casey Foundation, 2010).

Schools and districts across the United States are focusing their efforts on providing struggling readers the interventions they need through multi-tiered systems of support, often using the basic tenets of a Response to Intervention (RTI) framework (Berkeley, Bender, Peaster, & Saunders, 2009; Jenkins, Schiller, Blackorby, Thayer, & Tilly, 2013). An RTI framework relies on high-quality, core instruction with data-based decisions about intervention delivery for students not meeting predetermined achievement levels in core subjects (i.e., reading, language arts, and math). Evidence-based, targeted interventions in successive tiers create a framework for the provision of explicit instruction to students, with increased supports (e.g., reduced teacher/student ratio, increased dosage) based on each student's performance on continuous (e.g., weekly, biweekly) measures of progress (Jimerson, Burns, & Vanderheyden, 2007; Lembke, McMaster, & Stecker, 2010). Syntheses of reading research provide support for the use of interventions to improve the skills of K-12 students at risk for academic failure, as well as those already identified with learning disabilities (e.g., Flynn, Zheng, & Swanson, 2012; Scammacca et al., 2007; Swanson, 1999).

Though RTI provides a framework through which struggling students can access high-quality, targeted interventions, the implementation of multi-tiered support systems can be fraught with complications (Jimerson et al., 2007; Sullivan & Castro-Villarreal, 2013). Universal screening, tiered intervention delivery, and continuous monitoring of student progress are necessary procedures that require significant resources in time, materials, training, and personnel (Kovaleski, 2007). The provision of high-quality, evidence-based interventions targeting individual student needs can be particularly challenging, in that teachers, tutors, or volunteers must not only be available to conduct interventions, but must also be trained to implement interventions with a high degree of fidelity (Sullivan & Castro-Villarreal, 2013).

Many schools have turned to technology to remediate these resource challenges. For example, in 2008, the NCES reported that 98% of all public schools had available computers with Internet access. This percentage was stable across schools with large and small populations, community type (e.g., urban, rural) and percentage of students receiving free and reduced priced lunch. The ratio of students to computers in all public schools in 2008 was 3.1 to 1, with similar ratios reported for elementary and secondary schools. In a survey of teachers' use of educational technology, NCES (2010) reported that 69% of teachers used technology during instructional time in the classroom. When asked about software and Internet usage, 50% of teachers reported using drill and practice or tutorial programs with students, and 69% reported using technology to learn or practice basic skills during classroom instruction.

With increased availability, access, and use of technology in schools, the development of software and web-based programs designed to remediate basic academic skills presents an avenue of intervention delivery for students struggling with reading, writing, and mathematics. Rather than rely on trained personnel to deliver targeted interventions with fidelity to small groups of students, the use of such computer-assisted instruction (CAI) has the potential to provide targeted, *individualized* intervention for a large group of students at any given time. Given the availability of classroom computers, teachers could also potentially provide students with individualized intervention at times of the day that would not disrupt core learning opportunities. Indeed, CAI has been used effectively within core reading instruction and as a supplemental intervention delivery tool to boost the outcomes of struggling readers (e.g., Kim, Vaughn, Klingner, Woodruff, Reutebuch, & Kouzekanani, 2006; Macaruso & Rodman, 2009; Nicolson, Fawcett, & Nicolson, 2000; Wild, 2009).

The body of research investigating the effectiveness of CAI in reading is vast, spanning over five decades (Cheung & Slavin, 2012b). This research includes investigations of instructional technology at many levels, including educational technology supplemental to teacher instruction (e.g., interactive whiteboards), instructional technology supplanting teaching instruction (i.e., CAI), and interventions incorporating technology as a component of instruction. As multi-tiered systems of support are being incorporated into existing frameworks of many schools across the nation, syntheses of research on CAI used to deliver instruction in reading are needed to quantify the effect of this method of intervention delivery on reading outcomes. Given the responsibility of educators to conduct targeted interventions according to student need, an investigation of the effects of CAI at different grade levels and across different areas of reading will enhance an understanding of effective practices in intervention delivery.

Rationale for Study

Though primary research can shed light on the effectiveness of particular interventions, variability in research design within studies (e.g., treatment components, measured outcomes, data analysis) can convolute results across studies, limiting interpretations of evidence to support specific interventions or treatments. Research synthesis allows for the aggregation of primary research to help answer specific research questions regarding the effect a treatment (e.g., CAI) has on outcomes of interest (e.g., reading; Cooper, Hedges, & Valentine, 2009). Meta-analysis is a statistical technique of summarizing, integrating, and interpreting empirical research on a given topic (Denson & Seltzer, 2011; Lipsey & Wilson, 2001). This systematic method of locating, coding, and analyzing empirical research through an effect size metric enables researchers to draw quantitative conclusions about the cumulative state of research evidence (Cooper et al., 2009). Differences in results across studies can be analyzed by examining effects within subsets of studies (e.g., students in elementary or middle school), and regression techniques can be applied to help explain the variation in effects across studies, identifying variables that potentially moderate overall effects (Raudenbush & Bryk, 2002).

Previous meta-analyses of CAI have revealed small, positive effect sizes on overall academic and reading achievement. Investigations of CAI on overall academic achievement have revealed effects ranging from +0.33 to +0.40 (Christman & Badgett, 2003; Tamim, Bernard, Borokhovski, Abrami, and Schmid, 2011; Waxman, Lin, & Michko, 2013). Research disaggregating results in the area of reading have revealed effects ranging from +0.12 to +0.36 (Cheung & Slavin, 2012b; Kulik, 1994; Kulik & Kulik, 1991). Though these syntheses provide an evidence base supporting the use of technology to positively impact student achievement, operational definitions of CAI vary drastically among these syntheses and the studies they represent. This dissertation represents an expansion and refinement of previous meta-analyses to provide educators with information regarding the effectiveness of CAI, defined as an instructional delivery tool, and the factors that contribute to those effects. An analysis of research quality will help inform future research on CAI in reading and provide information for educators seeking CAI with high-quality research supporting evidence of improved outcomes.

School levels. Technology has changed educational practices from preschool through postsecondary programs. Given the increase in the use of RTI frameworks in schools across the country, this meta-analysis will remain focused on students in preschool through high school. By including preschool participants, the effects of CAI used in beginning reading instruction can be assessed. Including participants up to 12th grade allows for an assessment of remedial reading programs with older students. Disaggregation of results across school levels will provide additional information for educators at each level of instruction to make evidence-based decisions when choosing effective interventions. Information on the use of CAI to improve reading skills in adults after high school is beyond the scope of this synthesis.

Areas of reading. In 2000, the National Reading Panel (NRP) identified five areas that are essential to overall reading development: phonemic awareness, phonics, fluency, vocabulary, and comprehension. When students struggle with any of these five areas of reading, intervention should be targeted to that particular area and assessments of progress should align with the targeted need (Gersten et al., 2008; Kamil et al., 2008)). This review will disaggregate results across areas of reading to provide educators more information when assigning interventions to students struggling with particular aspects of reading.

Research quality. To determine future research paths, current research into the use of CAI in reading should be examined according to quality research design and implementation. Gersten, Fuchs, Compton, Coyne, Greenwood, and Innocenti (2005) developed a set of quality indicators for group experimental and quasi-experimental research in special education that can be used to evaluate the research at hand, and help scholars develop future studies with sound methodological design and implementation. The quality indicators suggested include (a) a concise description of participants, including abilities and disabilities, the comparability of groups, and the comparability of interventionists; (b) a detailed description of the intervention and its implementation, including the treatment and control conditions and fidelity measures; (c) multiple outcome measures timed appropriately; and (d) appropriate analysis and reporting of data, including effect sizes.

Research Questions

This dissertation will enhance the evidence base of CAI in reading by not only providing an updated estimation of overall effect, but also by using statistical analysis to investigate study characteristics contributing to those effects. The following research questions will guide this meta-analysis:

1. What is the mean effect of CAI on reading outcomes for students in preschool through high school and what is the evidence of variability among those effects?
2. What are the differential effects of CAI on reading outcomes across school levels (i.e., preschool/kindergarten, elementary, middle, and high school)?
3. What are the differential effects of CAI on reading outcomes across the five areas of reading (phonemic awareness, phonics, fluency, vocabulary, comprehension)?
4. What study characteristics contribute to the effects of CAI in reading (e.g., methodological characteristics, instructional components, research quality)?
5. What is the overall quality of the research according to the quality indicators proposed by Gersten et al. (2005)?

Definitions of Key Terms

The incorporation of technology in instruction can range from the simple use of calculators to the use of mobile, touch-screen devices to augment academic lessons. Given this variation in technology use, there is also variation in ways technology integration is defined in the literature (e.g., Cheung & Slavin, 2012b; Soe, Koki, & Chang, 2000). Existing literature often describes three distinct types of technology incorporated into instruction: (a) computer-assisted instruction, (b) assistive technology, and (c) educational technology.

Computer-assisted instruction (CAI). The definition of CAI in the context of this study is modeled after Moreno and Mayer's (2007) discussion of "interactive multimodal learning environments," (p. 310) which consist of a two-way interaction between the learner and the learning environment (i.e., computer or handheld device). Instruction is provided via the application rather than a teacher, volunteer, or classroom aid, and the application is both reliant on and responsive to the learner's actions. In the context of this meta-analysis, CAI is used with individuals or pairs of students, and a distinction is drawn between technology used as a part of core instruction (i.e., integrated in core literacy) and technology used for intervention (i.e., supplemental to core literacy).

Assistive technology (AT). The Individuals with Disabilities Education Act (IDEA, 2004) defines an AT device as "any item, piece of equipment, or product system…that is used to increase, maintain, or improve functional capabilities of a child with a disability" (20 U.S.C. 1401(1)). The purpose of AT is to allow students greater access to both information and learning (Boone & Higgins, 2007). McKenna and

Walpole (2007) define AT in the context of reading as, "any digital application that enables a user to comprehend text by supporting one or more components of the reading process" (p. 140; e.g., screen readers, text-to-speech devices, supported text through hyperlinks). The primary distinction between AT and CAI in this dissertation is the lack of application-led instruction in AT.

Educational technology (ET). ET represents the broad use of technology in the learning environment. Cheung and Slavin (2012a) define ET as "a variety of electronic tools and applications that help deliver learning content and support the learning process" (p. 4). ET can include both CAI and AT, as well as interactive whiteboards, LCD projectors, video-based instruction, and the internet (Cheung & Slavin, 2012b). The primary distinction between ET and CAI in this review is that ET can be used to augment teacher-led instruction (e.g., whole-class instruction using video anchors), whereas CAI relies exclusively on application-led instruction provided to individuals or pairs of students.

Summary and Contribution

Children who struggle with reading in any grade can benefit from interventions targeting specific skill deficits. Resource constraints can prevent schools from providing necessary interventions to students demonstrating academic challenges (Fletcher & Vaughn, 2009; King Thorius, Maxcy, Macy, & Cox, 2014). An instructional solution such as CAI may ensure more students receive individualized, targeted interventions when personnel resources are constrained. Previous syntheses on the use of CAI to improve outcomes in reading have revealed small, positive effects. This meta-analysis synthesizes primary research conducted since 2000 to summarize the effectiveness of this method of instructional delivery, to analyze factors contributing to potential effectiveness, and to examine the quality of available research on CAI in reading.

Structure of this Dissertation

This paper provides the findings of a quantitative synthesis of primary studies investigating CAI in reading. Chapter 2 provides the theoretical foundations guiding this meta-analysis and a literature review of previous narrative and quantitative syntheses of CAI in reading. Chapter 3 describes the methods used to search for relevant studies, code relevant variables, extract effects, and combine effect sizes to answer the research questions. Chapter 4 provides a description of the search process and results of all analyses. Finally, Chapter 5 places the findings within the context of theory and practice. Limitations of this study will be discussed, suggestions for educators looking to use CAI in their classrooms will be provided, and suggestions for future research examining the impact of CAI on reading outcomes will be presented.

Chapter 2

LITERATURE REVIEW

A systematic literature review not only helps to situate a primary study and its findings in the base of literature on a particular topic, it also helps to inform the path of future research. Given the nature of this paper, a thorough examination of previous narrative and quantitative syntheses of CAI in reading helps achieve these goals by positioning this meta-analysis in the context of other research syntheses and by informing the research design of this study. In this chapter, I first present theoretical models of reading development and multimedia learning, and discuss instructional design features derived from these models. Next, I briefly examine previous syntheses on the effects of CAI in reading (conducted previous to 2000). I then describe the method and results of an in-depth analysis of current narrative and meta-analytic syntheses (conducted since 2000). Finally, I discuss the implications of the findings and how those findings contribute to the design of this meta-analysis.

Theoretical Models and Instructional Design Features

Theories of reading development. A great deal of research has been conducted in an attempt to fully understand how children learn to read. Reading research highlights the importance of early literacy achievement for later academic success (Snow, Burns, & Griffin, 1998); students who do not master literacy skills in the early grades remain at-risk in both academics and social well-being throughout their schooling and into adulthood (Hernandez, 2011). Achieving proficiency in reading is a process that typically parallels a developmental trajectory to reach the end goal of gaining meaning from text (Kreskey, 2012).

Chall (1983) describes a six-stage model of reading development. In Stage 0 (e.g., pre-school), students are simply becoming acquainted with print. In Stage 1 (e.g., grades K-2), the child learns the alphabetic principle as well as decoding skills. In Stage 2 (e.g., grades 2-3) students practice decoding and begin to develop fluency and understanding of meaning in contextual reading. In Stage 3 (e.g., grades 4-8), instruction shifts from learning to read to reading to learn. Stage 4 readers (high school) must be able comprehend large quantities of material and report varying views. Stage 5 readers (college) are able to integrate and evaluate complex material potentially representing contradictory points of view.

Mastery of Stages 1 and 2 are necessary before readers are able to gain meaning from text, and poor readers are often "stuck" in these stages of reading development (Potter & Wamre, 1990). These readers require effective intervention to target particular stages of reading, such as phonemic awareness instruction for readers in Stage 1 and decoding instruction for readers in Stage 2. Ehri (2005) emphasizes that the "stages" described by Chall are actually "phases." Rather than completely mastering one level before moving to another, a child moves within phases, drawing on previous phases and components of more advanced phases to read and eventually understand the printed word.

Cognitive Load Theory. The developmental theories of Chall and Ehri emphasize the importance of early skill development; however, the theories are vague in relation to acquiring and teaching broad skills in reading comprehension. Cognitive Load Theory (CLT; e.g., van Merrienboer & Sweller, 2005) helps contribute to an understanding of reading development throughout the continuum. CLT posits that learning conditions are optimal when the amount of new information being learned does not overwhelm the capacity of working memory. When familiar material stored in schemas can be accessed, routines for processing information or performing actions can be accessed automatically, thus freeing up working memory capacity (Chandler & Sweller, 1991). In reading, a learner constructs increasingly complex schemas for different reading skills (e.g., decoding) by integrating previously mastered schemas for more basic skills (e.g., phonemic awareness), as is described in Chall and Ehri's developmental models of reading.

Three types of cognitive load are typically managed in instruction designed according to CLT: (a) intrinsic, (b) extraneous, and (c) germane (van Merrienboer & Sweller, 2005). Intrinsic load is imposed by the information to be learned. Extraneous load is imposed by activities and information that do not contribute to the process of creating or learning schemas. Germane load is imposed by the activities required to construct or master new schemas. CLT suggests that instruction be designed to limit extraneous load and redirect the learner's attention to cognitive processes relevant to schema construction (Sweller, van Merrienboer, & Paas, 1998). The application of CLT to reading development suggests that some students may need to be taught basic skills (e.g., decoding) separately from more complex skills (e.g., comprehension) to ensure that working memory demands are not overloaded (Kreskey, 2012). Explicit instruction can aid in the construction of schemas across all areas of reading, thus eventually leading students to the end goal of gaining meaning from text (Flynn et al., 2012)

Intervening with struggling readers. Syntheses of research provide support for the use of intervention in reading to improve the skills of K-12 students at risk for academic failure, as well as those already identified with learning disabilities (e.g., Flynn, Zheng, & Swanson, 2012; Scammacca et al., 2007; Swanson, 1999). Much of this work has focused on the use of *explicit instruction*, or instruction that targets specific skills and subskills, sequentially reviews previously learned materials, models new material to be learned, and provides guided and independent practice with corrective and elaborate feedback (Hall, 2002; Rupley, Blair, & Nichols, 2009). In a meta-analysis of studies on the use of explicit instruction in reading, Adams and Engelmann (1996) found a mean effect size of .75, which is substantial in educational research.

Areas of reading. The National Reading Panel (NRP; 2000) described five areas of reading that should be addressed in explicit reading instruction: phonemic awareness, phonics, fluency, vocabulary, and comprehension. *Phonemic awareness* is knowledge about the units of sounds in language (Foorman & Torgesen, 2001). Activities in phonemic awareness often involve orally manipulating individual sounds in words. *Phonics* is knowledge of letter-sound correspondence. Phonics instruction builds on phonemic awareness by helping children understand that the sounds they hear in words are made up of specific letters. *Fluency* is the ability to accurately and quickly decode text in order to facilitate comprehension. *Vocabulary* involves the ability to understand and articulate the meaning of words. The NRP (2000) suggests that vocabulary should be taught directly by teaching meaning of specific words, and indirectly by engaging children in rich verbal interactions with people who are linguistically proficient. *Comprehension* is often thought of as the end goal of reading development, and involves the ability to make literal and inferential meaning from what is read. The NRP (2000) suggests that reading comprehension is an active process that involves integration of background knowledge and expertise to construct mental representations of what is read.

Design components of explicit instruction. Each of these five areas of reading can be taught using six distinct design components of an explicit instruction model (Hall, 2002): (a) big ideas, (b) conspicuous strategies, (c) mediated scaffolding, (d) strategic integration, (e) judicious review, and (f) primed background knowledge. By teaching *big ideas,* teachers can direct learners to the overall objectives of a learning task, thereby orienting their learning and facilitating a broader acquisition of knowledge. *Conspicuous strategies* involve explicit instruction in specific strategies (e.g., mnemonic devices) to help complete tasks or solve complex problems. *Mediated scaffolding* involves temporary support and guidance during initial learning to reduce task complexity and increase successful task completion. *Strategic integration* allows teachers to combine essential information with big ideas in multiple contexts in an effort to enhance generalization of learning beyond the classroom. When teachers employ *judicious review* they promote the maintenance of conceptual and procedural knowledge. By *priming background knowledge*, teachers can enhance learning by situating the lesson within the knowledge the learner brings to the task and the accuracy of that knowledge (Hall, 2002).

Delivery components of explicit instruction. The design components of explicit instruction are delivered through the following five instructional delivery features (Hall 2002): (a) frequent student response, (b) appropriate pacing, (c) adequate processing time, (d) response monitoring, and (e) feedback for correct and incorrect responses. By allowing *frequent student response*, teachers can keep instruction highly interactive and engaging while also providing students time to practice new learning. *Appropriate pacing* should be determined based on task complexity, background knowledge, and individual student differences. *Adequate processing time*, much like pacing, varies dependent on the student's understanding of the material and the task the student is expected to perform. All teachers must constantly *monitor responses* to determine student mastery of skills during instruction. Key to effective instructional delivery is the provision of *feedback for correct and incorrect responses*. This feedback should be specific and instructional, and should not interfere with the pacing of the instruction (Hall, 2002).

Cognitive theory of multimedia learning. Mayer's (2005) cognitive theory of multimedia learning provides the lens through which CAI is defined in this study and also provides a guide toward analysis of effective instructional components of CAI. This theory is based on CLT and describes three cognitive principles of learning: (a) the human information processing system includes dual channels (i.e., audio and visual); (b) these channels have limited processing capacity; and (c) active learning consists of active processing during learning. As described in CLT, stimuli representing extraneous load has the potential to strain cognitive resources (Kennedy & Deshler, 2010). Mayer's (2005) theory recognizes the need for active learning through multiple modes of representation (i.e., audio and visual) while optimizing instructional design to limit extraneous processing and encourage schema production (Kennedy & Deshler, 2010; Moreno & Mayer, 2007).

Drawing from this theory and extensive research, Moreno and Mayer (2007) have identified five instructional design principles essential for CAI: (a) guided activity, (b) reflection, (c) feedback, (d) pacing, and (e) pretraining. *Guided activity* involves interaction with a pedagogical agent (e.g., a cartoon figure or narrator) to help prompt the organization and integration of new information, thus encouraging cognitive processing and schema construction. When CAI incorporates *reflection*, students have the opportunity to analyze their responses and integrate new information. *Feedback* is essential, and explanatory feedback, rather than corrective, can provide students appropriate schemas to address misconceptions. When students control the *pace* of their instruction, smaller chunks of information can be processed in working memory at one time. *Pretraining* involves priming background knowledge, which helps students integrate new information with prior knowledge. Research into each principle has been conducted across a wide range of content areas (e.g., science, math, reading) and with a wide range of populations (e.g., elementary students, college students, and preservice teachers; see Moreno & Mayer, 2007). The alignment of Moreno and Mayer's (2007) recommendations for multimedia learning with components of explicit instruction recommended in intervention literature (Hall, 2002; Swanson, 1999) is summarized in Table 1.

Table 1

Alignment of CAI Design Principles and Explicit Instruction

CAI Component	Description	Alignment with Components of Explicit Instruction
Guided Activity	Pedagogical agent to guide processing – in CAI this agent acts as the teacher	Big Ideas Conspicuous strategies Feedback
Reflection	Analyze responses and integrate new information	Judicious Review Strategic Integration
Feedback	Corrective or elaborate feedback based on student responses	Feedback for Correct and Incorrect Responses Response Monitoring
Pacing	Student controls the pace of instruction and activities	Mediated Scaffolding Appropriate Pacing Adequate Processing Time
Pretraining	Priming background knowledge to aid in the construction of new schemas	Judicious Review Primed Background Knowledge

Though there is limited research investigating the effectiveness of Moreno and Mayer's (2007) design features in reading, several scholars have offered similar recommendations worth noting (i.e., Bishop & Santoro, 2006; Kennedy & Deshler, 2010; Smith & Okolo, 2010). Smith and Okolo (2010) investigated commercially available applications for evidence-based features of explicit instruction. Their recommendations aligned with those of Moreno and Mayer (2007), adding the need for conspicuous strategies (e.g., a set of steps to solve a given problem) within the instructional design of CAI. As noted in Table 1, conspicuous strategies is a component recommended in explicit instruction design, and could be addressed through the guided activity feature suggested by Moreno and Mayer (2007).

Bishop and Santoro (2006) analyzed commercially available software for design features related to program interface, content, and instruction. They focused their recommendations on the balance between these three aspects of CAI in an effort to reduce cognitive load and allow for optimal cognitive processing. Bishop and Santoro (2006) emphasize the need for continuous assessment of learner response and adjustment of the program accordingly. They also suggest that the results of assessments be tracked, allowing educators to determine the effectiveness of CAI for each particular student. Santoro and Bishop (2010) examined 31 commercially available beginning reading programs for their proposed recommendations (Bishop & Santoro, 2006). They found limited use of these recommended practices among the programs they investigated. Their study highlights the need for a focus on instructional design in CAI research in order to inform best practice for educators choosing available programs and for developers creating and updating CAI applications.

Previous Syntheses of Research on CAI

For more than four decades, researchers in the field of education have been investigating the use of computers in reading instruction. One of the first programs reported in the literature, the Stanford CAI System, had the capability of making real-time decisions while providing instructional branching contingent on student response. This progressively designed program was found to be an effective supplement to teacher-led reading instruction (Fletcher & Atkinson, 1972). Advances in technology throughout the decades have led to more sophisticated program designs incorporating academic and social behaviors.

Meta-analyses of CAI have been conducted since the 1980s and show consistency in outcomes. Kulik and Kulik (1991) conducted one of the earliest syntheses of CAI research, including various academic outcomes and participants ranging from kindergarten to adulthood. They found a mean effect of +0.30 among 254 studies. Incorporating these results into a review of other meta-analyses of CAI, Kulik (1994) reported a median effect of +0.36 among syntheses. Cheung and Slavin (2012b) also reviewed previous syntheses on CAI and found a range in effect from +0.12 to +0.18 among students in grades K-12 (see Becker, 1992; Fletcher-Finn & Gravatt, 1995; Ouyang, 1993). These effect sizes have been interpreted as small in social science research (Cohen, 1988); however their positive nature suggests that increases in academic achievement can occur through the use of CAI. No syntheses published prior to 2000 specifically focusing on reading CAI could be located for this review.

Several meta-analyses of CAI research conducted since 2000 have focused on overall academic outcomes without disaggregating effects on reading. The effect sizes reported in these studies are consistent with effects found in the previous two decades. For example, Christmann and Badgett (2003) found a mean effect of +0.34 among 39 studies of students in grades K-6. Similarly, Waxman et al. (2003) found a mean effect of +0.40 among 42 studies examining the addition of technology to learning environments. Tamim et al. (2011) found a mean effect of +0.33 in a second-order meta-analysis analyzing effects reported in 25 meta-analyses across 40 years of research.

The consistency of effects across syntheses of CAI indicates that this method of instruction can produce positive academic outcomes for K-12 students. Since 2000, syntheses of CAI research in particular academic areas (e.g., reading, math) have begun to be published. Though each synthesis of CAI reading research provides an indication of overall effects on particular reading outcomes, a comprehensive literature review of current syntheses (published since 2000), much like the secondary analyses conducted by Kulik (1994) and Tamim et al., (2011), can shed light on effects across a large number of primary studies, and can inform directions for future primary research. The results of this literature review will also shed light on gaps in meta-analytic research, thus informing the work represented in this meta-analysis. The following questions guided this literature review: (a) What do recent narrative and meta-analytic syntheses report about the effectiveness of CAI in reading for students in preschool through twelfth grade? (b) What instructional design features (e.g., intensity, level of feedback) of available programs have been shown to contribute to these effects?

Literature Review Method

Search Procedure

Systematic searches were conducted in four databases: Academic Search Premiere, Education Full Text, ERIC, and PsycINFO. Available papers included both peer-reviewed journal articles and non-peer-reviewed papers published elsewhere (i.e., meta-analyses published online; Cheung & Slavin, 2012b; Soe, Koki, & Chang, 2000). Each database was searched using the terms (reading or literacy) and (technology or computer or computer-based or computerized or computer-assisted or software) and (instruction or intervention or education or special education or RTI or tier). The abstracts of the resulting articles were reviewed to identify studies for possible inclusion. The reference lists for studies meeting the criteria listed below were also reviewed to identify additional articles for possible inclusion.

Inclusion Criteria

Articles were included if they provided a narrative or quantitative synthesis of primary CAI studies in reading. CAI syntheses involving students in preschool through twelfth grade were included in this review, regardless of disability status. Articles were included if the primary outcome measures involved one or more areas of reading defined by the National Reading Panel (NRP, 2000; i.e., phonemic awareness, phonics, fluency, vocabulary, comprehension). Syntheses investigating studies using measures in reading and another subject area (e.g., writing) were included, as long as reading results were disaggregated in the findings. Articles were excluded if they reported results of one primary study or provided commentary on CAI without synthesizing primary research. Syntheses published before 2000 were also excluded in order to examine knowledge that is relevant to currently available CAI technology; however, most syntheses meeting inclusion criteria for this review included primary articles published before 2000.

Research Characteristics

In addition to examining reported effects, I extracted several research characteristics to examine the constitution of available research. Variables examined (see Table 2, p. 40), included the publication year of each synthesis, the type of technology examined (i.e., CAI, AT, or ET), outcomes measured, student characteristics, the number of cited studies, the number of studies unique to each synthesis, the median year represented among cited studies, overall findings, reported effect sizes, and reported confidence intervals. The number of unique studies was computed by coding each citation separately and determining the amount of overlap with previously published syntheses. Descriptions of instructional design features and technology type were examined to answer the second research question of this review, and available information is summarized in Table 3 (p. 42).

Reporting Overall Effects

Effect sizes provided within the meta-analyses were reported as written, with one exception. Soe et al. (2000) reported an *r*-effect size, while the other meta-analyses reported either Cohen's *d* or Hedges *g*. The *r*-effect in Soe et al. (2000) was converted into Cohen's *d* using an online conversion calculator based on effect size calculations suggested by Lipsey and Wilson (2000; http://www.campbellcollaboration.org/resources/effect_size_input.php) This conversion to a common effect helps ease the interpretation and comparability of overall effects. Several syntheses reported one overall effect size for all given studies, while others (i.e., Cheung & Slavin, 2012a; 2012b; Kulik, 2003) disaggregated effect sizes for CAI as defined in this paper from overall effects of educational technology. The number of effect sizes contributing to reported mean effects (*k*) is provided for syntheses including more than one effect per study. In an effort to quantitatively aggregate the results found across meta-analyses, the mean, median, standard deviation, and frequency distribution of effect sizes were computed using SPSS™ (Version 20) statistical software.

Literature Synthesis

Effectiveness of CAI in Reading

A total of 13 syntheses (four narrative, nine meta-analytic) met criteria for the first and second research questions. Reports were published between 2000 and 2012 and contained primary articles published in years ranging from 1979 through 2011. Table 2 provides a summary of synthesis characteristics and results. Though none of the included syntheses delineated instructional design components of CAI contributing to effects, several articles described categories of technology that are summarized in Table 3. Three syntheses were not published in peer-reviewed journals (Cheung & Slavin, 2012b; Kulik, 2003; Soe, Koki, & Chang, 2000) and were located online (e.g., www.bestevidence.org). A total of 224 unique citations of primary studies were found within these 13 studies, 174 (78%) of which were reviewed in only one paper. The number of CAI studies reviewed in each synthesis ranged from 5-82, with a median of 18. In the remainder of this section, I will describe the results of narrative syntheses and meta-analyses separately, followed by an overall summary of effects.

Narrative reviews. Narrative reviews provide a qualitative means by which to examine primary research. Whereas meta-analyses use quantitative measures to independently compute and analyze effect sizes, narrative reviews rely on information reported within each primary study to analyze and evaluate research. A focus on in-depth descriptions of available literature allows for a rich discussion and critical analysis of study features (Cooper & Hedges, 2009).

Four narrative reviews met criteria for inclusion in this paper. Positive findings for reading outcomes were noted in all four reviews. Three reviews examined the effectiveness of CAI on the reading achievement of students with learning disabilities (LD; see Table 2). The report offered by the National Reading Panel (NRP; 2000) represented the only narrative review to include K-12 students with and without disabilities. The panel examined 21 experimental and quasi-experimental, peer-reviewed studies ranging from preschool through high school. Included studies were published between 1986 and 1996. The panel felt that the small sample of studies precluded meta-analysis, which was the method of synthesis used within several other areas in the overall report. A variety of outcomes were represented, including spelling (n = 1), vocabulary (n = 2), comprehension (n = 2), broad reading skills (n = 2), and the addition of a text to speech component to reading instruction (n = 6). The type of CAI examined was categorized as *replacement* (i.e., AT), described as using the computer to complete tasks that could be performed with paper and pencil, or *augmentation* (i.e., CAI), in which computers provide appropriate instruction based on student answers (see Table 3). All included studies within both categories reported positive outcomes. The panel suggested that future research focus on essential program components, the effective integration of CAI in reading, and the addition of word processing in reading instruction.

In the same year, Hall, Hughes, and Filbert (2000) published a review of CAI in reading for students with LD. Experimental and quasi-experimental, peer-reviewed studies (n = 17) published between 1980 and 1997 were included, none of which overlapped with studies examined by the NRP. In 13 of the studies, there was demonstrated improvement in reading decoding or reading comprehension. Studies investigated CAI using drill-and-practice procedures (*n* = 10) and strategy instruction (*n* = 6), with one study investigating simulation (see Table 3). Elaborate feedback (e.g., corrective feedback with additional practice) was cited as an effective component in seven studies, all of which reported statistically significant effects. Authors reported no apparent patterns to the type of CAI or area of reading instruction for the four studies with no between-groups differences. Given computer technology to target specific skills on an individual basis, Hall et al. (2000) posit that teacher to student ratios can be reduced, increasing opportunities for high quality, instructional interactions among teachers and their students.

MacArthur, Ferretti, Okolo, and Cavalier (2001) also published a narrative review examining the effects of technology in literacy among students with LD. Reading studies (*n* = 28) were examined for outcomes in word identification and text comprehension. Though the published dates of included reading studies (1985-1998) encompassed the interval of time represented in Hall et al. (2000), there were only two studies represented in both syntheses. The reason for such little overlap between these two studies is unclear, as they both had markedly similar purposes and search procedures. MacArthur’s study included five studies reviewed in the NRP report (2000), and 21 studies were newly synthesized. Results of MacArthur et al.’s (2001) review indicate that the bulk of research investigated CAI for word identification remediation. All studies reviewed supported the efficacy of CAI in improving phonological awareness and decoding. The provision of speech feedback and electronically enhanced text to improve comprehension demonstrated promise, but was cited as an area needing further research. These two types of technology fit the definition of AT in this paper, suggesting a lack of research on the effectiveness of application-led instruction (i.e., CAI) to improve comprehension.

The most recent narrative review of CAI in reading was conducted by Stetter & Hughes (2010). Similar to MacArthur et al. (2001), these authors focused their investigation on struggling readers and students with LD. The outcome of interest was the effect of CAI on reading comprehension. Studies (*n* = 27) from 1985 through 2009 were examined, with 12 studies also represented in MacArthur et al. (2001) and three studies in Hall et al. (2000).The authors stated that studies focusing on at-risk students were not included, though no discriminating definition between 'struggling' and 'at-risk' was provided. Nine studies were unique to this review, and 11 studies were published in 2000 or later. In this review, studies were placed in one of three categories, all of which fit the definition of AT in this paper: (a) computer-presented text vs. paper presentation; (b) computerized readers providing support for unknown words; and (c) programs providing supported text via hypertext. Studies examining software incorporating comprehension instruction were included in the literature search, but were not located for Stetter and Hughes's (2010) review.

Results indicated that the digital presentation of text made little difference for students over traditional paper presentation, and four of the seven studies examining computerized readers reported positive outcomes. Supported digital text was examined in 15 studies, 12 of which reported positive results. Only three of these studies reported significant differences in the area of comprehension. Conclusions from this review, much like those of MacArthur et al. (2001), point to the potential of AT for improving comprehension. Stetter and Hughes (2010) highlight the need for more research in this area, including the use of CAI to provide explicit instruction in comprehension, rather than relying on AT to provide text support.

Meta-analyses. Nine meta-analyses met criteria for inclusion in this review (see Table 2). Two meta-analyses included K-12 students with no specification as to academic challenges or disability status (Soe, Koki, & Chang, 2000; Cheung & Slavin, 2012a); the remaining meta-analyses included specific populations of students. Jitendra, Edwards, Sacks, and Jacobson (2004) published the only meta-analysis focusing on students with LD, and two syntheses with considerable overlap in studies (Slavin, Lake, Davis, & Madden, 2011; Cheung & Slavin, 2012b) examined the effects of CAI on the outcomes of struggling readers. Overall reported effect sizes (k = 12) ranged from -0.03-0.49, with a mean of +0.15 (SD = 0.13) and a median of +0.13. Strong, Torgerson, Torgerson, and Hulme’s (2011) meta-analysis of Fast ForWord represented the only synthesis to find negative effects. The remaining eight meta-analyses found small to moderate, positive effects (range = +0.06-0.49).

A technical report partially funded by the U.S. Department of Education was published in 2000 examining the effects of CAI on the reading achievement of students in grades K-12 (Soe, Koki, & Chang, 2000). Included studies (n = 17) were published between 1982 and 1997. Though this range overlaps with the published years covered by the NRP (2000), only two studies are represented in both papers. None of the studies in Hall et al. (2000) were included, and 13 of the studies cited were unique to this review.

Soe et al. (2000) defined CAI as learning from computers, or using the computer to transmit specific subject matter. They categorized CAI in three levels (see Table 3): (a) drill and practice – providing reinforcement with specific skills and supplying immediate feedback; (b) tutorial – clarification of concepts tailored to the individual; and (c) dialogue – allowing the student to take an active role in structuring the curriculum. Though 40 individual effects were reported among the 17 studies, these effects were combined to obtain one effect size per study. Studies with larger sample sizes received more weight when computing the overall effect (r = +0.13; d = +0.26), which was statistically significant. The authors did not separately analyze the contribution of the technology categories they mentioned in their report, nor did they discuss the areas of reading (e.g., comprehension, vocabulary) covered by the programs. A brief analysis of scatter plots, including sample size, treatment duration, and grade level revealed no particular study characteristics to systematically explain variation in effects. Only five of the 17 studies were published in peer-reviewed journals, though the contribution of publication type on overall effect was not addressed.

A team of researchers from the University of Amsterdam (Blok, Oostdam, Otter, & Overmaat, 2002) expanded and updated Soe et al.'s (2000) findings by synthesizing the effects of CAI on beginning reading skills (i.e., phonological and phonemic awareness) across 42 studies published between 1990 and 2000. Over 70% (n = 30) of the studies analyzed were unique to this publication. The population of students ranged in age from 5-12.5 years and the authors placed programs into six categories (see Table 3): (a) phonological awareness training (n = 15); (b) word reading with speech feedback (n =

14), (c) flashed word reading (e.g., digital flash cards; n = 13); (d) text reading with speech feedback (n = 12); (e) reading while listening (n = 2); and (f) mixed programs (n = 18). Outcome variables included (a) phonological skills, (b) letter identification, (c) word identification accuracy, (d) word identification speed, (e) oral text reading accuracy, (f) oral text reading speed, and (g) omnibus reading or mixed measures.

In the analysis, study characteristics were entered in progressive steps to look for explained variance in overall effects. The strongest associations were effect size at pre-test, with higher effect sizes for experimental groups displaying advantage at pre-test, and language of instruction, with higher effect sizes for programs provided in English. This two-predictor model reduced the variability of study outcomes by 61% and led to an overall mean effect of +0.19 (k = 50), which was statistically significant and similar to the findings of Soe et al. (2000). An analysis of studies investigating English-speaking programs alone yielded a mean effect size of +0.50. The authors did not discuss effect sizes within categories of programs or among types of outcomes represented, nor did they address the potential dependence of effects resulting from the combination of several effect size estimates from each primary study (Cooper et al., 2009). Though non-peer-reviewed studies were included in this synthesis, an analysis of the contribution of publication type was not conducted.

In a report prepared for a non-profit research firm, Kulik (2003) investigated the effects of instructional technology on the outcomes of reading and mathematics in elementary and secondary schools. Kulik (2003) categorized the technology represented in 27 studies as (a) integrated learning systems (ILS); (b) writing-based reading programs (i.e., *Writing to Read*); and (c) reading management programs (i.e., *Accelerated Reader*). None of these studies had been previously synthesized. ILS were defined as software providing sequential tutorial instruction while keeping records of student progress. The writing-based reading programs and reading management programs contained technological components that were part of a larger core instructional program; thus, the effects reported for these categories are not included in the overall mean effect in this paper. The synthesis of studies investigating ILS (*n* = 9) ranged in year from 1991 to 1996, and Kulik (2003) found little evidence for the effectiveness of these systems (ES = +0.06) among students in grades 2-8. All nine studies investigated outcomes in both math and reading, and effects were found to be higher for mathematics (+0.17). The effects for writing-based reading programs (range = +0.25-0.84) and reading management programs (+0.43) proved to be much larger than ILS, though the number of studies used to compute these effects ranged from two to six studies. The inclusion of both math and reading during the ILS interventions, as well as the inclusion of four dissertations among these nine effects, should be considered when attempting to generalize these results to other CAI programs.

Two meta-analyses have investigated the effects of CAI on the reading outcomes of older students. Jitendra et al. (2004) investigated research on effective vocabulary instruction for students with LD in grades 4-12. Out of the 27 peer-reviewed studies investigated, interventions using CAI encompassed six studies, with two of those providing information used to calculate a mean effect size (+0.16). All of the studies investigated were unique to this synthesis, though the range of publication year (1983-

1996) was similar to previously described syntheses. Among the six studies examined, four reported positive effects. One of the two studies contributing to the mean effect examined the use of video anchors to activate students' prior knowledge (Koury, 1996). A large negative effect (-1.79) was reported in this study, possibly due to the comparison of students with LD with general education students also receiving treatment. The second study contributing to the mean effect reported a very large positive effect (+2.22) of CAI on experimental measures of vocabulary (Horton, Lovitt, & Givens, 1988). The fact that only two studies with extremely wide variation in effect contributed to the quantitative analysis should be considered when interpreting this overall mean effect. Furthermore, Koury (1996) investigated video anchors, technology that does not provide instruction, thus failing to meet this paper's definition of CAI. This synthesis represents the only one reported in this paper to describe the effects of vocabulary instruction; given the limitations of this synthesis, more research into the effectiveness of vocabulary CAI should be conducted.

Moran, Ferdig, Pearson, Wardrop, and Blomeyer (2008) also investigated the effects of CAI on the reading performance of older students (grades 6-8) through an investigation focused on reading comprehension. The 20 included studies ranged in publication year from 1988 to 2005, with 16 studies (80%) published in 2000 or later. Only studies published in peer-reviewed journals were included. Of the 20 studies, 18 had never been included in a previous synthesis. The overall weighted effect (k = 89) was +0.49 and statistically significant. This effect size represented the highest among the studies examined in this review. The type of technology analyzed among these studies was separated into three categories (see Table 3): (a) commercial (e.g., comprehensive programs sold to the general public); (b) researcher-designed learning environments; or (c) delivery systems built upon research-based principles (e.g., electronic text with an available dictionary).

Results indicated that researcher-designed learning environments led to a higher mean effect (+1.20) than commercial software (+ 0.28) or evidence-based delivery systems (+0.34), with statistically significant differences among the categories. Several research design characteristics, including larger sample sizes (i.e., >30), researcher-designed assessments, general education participants, and studies using a correlated design all yielded effect sizes that were significantly higher than studies with other characteristics. Part of the analysis compared meaning-focused interventions (e.g., comprehension, vocabulary, and metacognition) with code-focused interventions (e.g., phonics, phonemic awareness, and fluency); however, there was no mean effect size favoring one reading emphasis over another. Results indicated that CAI may be more effective for general education students than for special education students; however, the authors briefly cite issues of engagement and levels of support as explanations of this finding. Further analysis of study quality and measurement techniques within the included primary studies would help shed light on this result. This meta-analysis allows the first glimpse into the types of CAI and research characteristics that contribute to overall effects; however, the inclusion of dependent effect sizes in the overall analysis should be considered when interpreting results.

A unique meta-analysis located for this synthesis analyzed the effects of one widely available program, Fast ForWord (FFW), on single-word reading, passage comprehension, receptive vocabulary, and expressive vocabulary (Strong et al., 2011). Only six peer-reviewed studies were analyzed, one of which was represented in Moran et al. (2008). All of the studies analyzed were published after 2000. Comparisons with untreated controls (+0.08) and controls using other programs (-0.03) revealed effect sizes with confidence intervals including zero. Authors cited no evidence to show that FFW was effective for improving vocabulary or overall reading skills, though the small sample of studies analyzed should be taken into account when considering the implications of these findings. In fact, in two intervention reports synthesizing hundreds of studies on FFW, the What Works Clearinghouse (WWC, 2010; 2013) found positive effects on alphabetics for beginning readers and potentially positive effects on reading fluency and comprehension for adolescent readers. Several of the studies reviewed in Strong et al. (2008) were included in the WWC analysis, though the WWC included technical reports and non-peer-reviewed studies in their final report.

Since 2011, a group of researchers have conducted three meta-analyses investigating the effects of CAI on reading achievement (Cheung & Slavin, 2012a, 2012b; Slavin, et al., 2011), with considerable overlap in represented primary studies among the three syntheses, but very little overlap with previously conducted narrative reviews and meta-analyses. In fact, out of the 88 studies cited among these three syntheses, only eight had been previously reviewed.

In a meta-analysis examining the effects of reading interventions for struggling K-12 students, Slavin et al. (2011) investigated the following five methods of instruction: (a) one-to-one tutoring with a teacher; (b) one-to-one tutoring with a paraprofessional or volunteer; (c) small group tutorials; (d) classroom instructional procedures (e.g., Peer-Assisted Learning Strategies, Direct Instruction); (e) Success for All, and (f) instructional technology. A weighted mean effect of +0.09 (k = 14) across 12 studies of technology was reported, compared to mean effect sizes ranging from +0.16 for one-on-one tutoring with volunteers to +0.56 for classroom instructional process, including Success for All (ES = +0.55). Outcomes examined included only measures of overall reading performance; studies using measures inherent to treatment (e.g., researcher-developed assessments) or examining specific areas of reading (e.g., phonemic awareness) were excluded from this review. Additionally, inclusion criteria required comparison groups to receive standard or alternative treatments, treatment duration of at least 12 weeks, and the inclusion of at least two teachers in each treatment group to control for teacher effects. Both peer-reviewed and non-peer-reviewed studies were included in the analysis, though an investigation of this contribution to the overall effect was not conducted.

A closer examination of cited CAI studies in Slavin et al. (2011) reveals a range in year from 1978 to 2009, with 10 of the 12 studies published prior to 2000. The three smallest effect sizes were found among three of four large scale designs included in the analysis (Campuzano et al., 2009; Dynarski et al., 2007; Rouse & Krueger, 2004). These three studies contributed considerably more weight to the overall mean and may have artificially deflated results. The experimental control of Campuzano et al. (2009) and

Dynarski et al. (2007) is questionable (e.g., quality of the programs examined, fidelity of program use), and the low effect of the large-scale Rouse & Krueger (2004) study is reflected in Strong et al.'s meta-analysis of FFW. Moderating effects of specific program components or areas of reading instruction were not reported in this synthesis. The small sample size included in this analysis, as well as the stringent inclusion/exclusion criteria mentioned above limits the generalizability of findings related to CAI in reading.

Cheung and Slavin (2012a) expanded their investigation of the effects of technology on the outcomes of struggling readers in a non-peer-reviewed study found in the Best Evidence Encyclopedia, a free online website dedicated to evidence-based practices in education (http://www.bestevidence.org). In this synthesis, the authors included studies of educational technology (ET), defined as electronic tools to deliver learning content and support the learning process. Four types of ET were categorized (see Table 3): (a) traditional, supplemental CAI (e.g., drill-and-practice, self-tutorial materials); (b) comprehensive models (e.g., CAI alongside core reading approaches; *Read 180*); (c) small-group integrated models (e.g., *Lindamood Phoneme Sequence*); and (d) FFW, as described in Strong et al. (2011). For the purposes of this paper, only supplemental programs and FFW were considered CAI. Both published and unpublished studies (*n* = 18) included students in grades K-6 and ranged in year from 1980 to 2012. Though inclusion/exclusion criteria were almost identical to Slavin et al. (2011), three articles in the previous synthesis were not included in Cheung & Slavin (2012a), and the more recent synthesis added eight articles that had not previously been synthesized. Similar to Slavin et al. (2011), studies using measures of skills inherent to the intervention (e.g., researcher-developed assessments) and measures of specific reading skills were excluded.

The overall weighted mean effect found by Cheung and Slavin (2012a; +0.14; *k* = 20) was higher than the previous estimate and statistically significant. A more thorough analysis found that small-group integrated applications had the largest effect (+0.32; *k* = 3), followed by supplemental programs (+0.18; *k* = 12). Comprehensive models (+0.04; *k* = 3) and FFW (+0.06; *k* = 2) demonstrated smaller effects. The effect size for supplemental programs was the only category demonstrating a mean effect significantly different from zero; it also represented the category with the largest number of studies analyzed. Results indicated that technology may be more effective for younger students (K-3; ES = +0.36; *k* = 8) than older students (4-6; ES = +0.07; *k* = 10). Consistent with previous syntheses, studies with large samples (>250) using a randomized design demonstrated the smallest effects (e.g., Campuzano et al., 2009; Dynarski et al., 2007). Effect sizes for studies from published journals were significantly larger than those from unpublished sources.

The effect size of +0.18 for supplemental CAI programs found by Cheung and Slavin (2012a) is slightly larger than the median effect found in this synthesis, though the small sample of studies analyzed (*k* = 12) should be considered, as well as the weighted contribution of two studies with the lowest effects and the largest sample sizes (Campuzano et al., 2009; Dynarski et al., 2007). The small effect of FFW is not surprising, given the inclusion of Rouse and Krueger (2004), the large-scale study analyzed in Strong et al. (2011). In fact, Slavin et al. (2011) included this study in their analysis of instructional technology while Cheung and Slavin (2012a) placed the effects of FFW in a separate category of analysis, which helps explain the lower effect size found in Slavin et al. (2011) compared to the mean effect of CAI reported Cheung and Slavin (2012a).

Cheung and Slavin (2012b) published a larger meta-analysis examining the effectiveness of ET on the reading outcomes of students in grades K-12 with and without academic challenges. Inclusion/exclusion criteria were virtually identical to Slavin et al. (2011) and Cheung & Slavin (2012a). Only comprehensive reading outcomes were included; studies with primary measures of phonemic awareness, oral vocabulary or writing were excluded. A total of 82 published and unpublished studies available from 1978 to 2010 were examined, with almost 60% of studies appearing after 2000. Eight studies overlap with previously mentioned narrative and quantitative syntheses conducted by authors outside of this research team, including two represented in the NRP report (2000), five represented in Kulik (2003) and the large scale FFW study (Rouse & Krueger, 2004) mentioned in both Moran (2005) and Strong et al. (2010). Authors categorized interventions (see Table 3) as (a) computer-managed learning (e.g., *Accelerated Reader*); (b) innovative technology applications (e.g., *FFW*); (c) comprehensive models (e.g., *Read 180*); and (d) supplemental technology. The categories of innovative technology applications and supplemental technology fit the definition of CAI provided in this paper and were thus included in the overall mean effect of this synthesis.

An overall weighted mean effect of +0.16 indicated a small, positive effect of ET on overall reading achievement consistent with other syntheses (Blok et al., 2002; Cheung & Slavin, 2012a; Jitendra et al., 2004; Soe et al., 2000). Studies utilizing comprehensive models (*k* = 18) produced the largest effects (+0.28), followed by computer managed learning (*k* = 4; ES = +0.19) and innovative technology applications (*k* = 6; ES = +0.18). Supplemental technology, or CAI (*k* = 57), was found to have the lowest overall effect (+0.11). All four mean effects were statistically significant, though between-group differences were not significant. There was some indication that ET may be more beneficial for English language learners and students with academic challenges. Other variables analyzed (e.g., program intensity, fidelity of implementation, SES, gender) did not produce significant differences in effect. Consistent with previous research, primary studies with large, randomized samples (e.g., Dynarski et al., 2007) showed smaller effects, and effects reported in published studies (*k* = 21; ES = +0.25) were significantly higher than effects reported in unpublished studies (*k* = 63; ES = +0.14). The overall effect for supplemental CAI in this synthesis was obtained from the largest number of studies included in the meta-analyses reported here. A closer examination of variables moderating this overall effect would be helpful to research investigating effective practices within CAI. Furthermore, information regarding the differences between innovative technology applications and supplemental CAI was lacking, and the higher mean effect for the former category suggests variation that should be explored.

Table 2

Characteristics of CAI Syntheses

Synthesis Type	Study	Tech Type	Outcomes	Participants	Studies Cited	Unique Studies	Median Year	Descriptive Findings	ES	Confidence Interval
Narrative	NRP, 2000	CAI	Overall	Grades K-12	21	10 (48%)	1989	All studies reported positive outcomes	*	*
	Hall et al., 2000	CAI	Overall	Grades K-12 with LD	15	10 (67%)	1987	Demonstrated improvement in 13 studies	*	*
	MacArthur et al., 2001	CAI, AT	Word ID, Comp	Grades K-12 with disabilities/ at-risk	28	5 (18%)	1992	All studies report improved phonological awareness and decoding	*	*
	Stetter & Hughes, 2010	ET, AT	Comp	K-12 with LD; at-risk not included	27	9 (33%)	1995	Providing additional text supports may improve comprehension	*	*
Meta-analysis/ Best evidence synthesis	Soe et al., 2000	CAI	Overall, Comp	Grades K-12	17	13 (76%)	1993	Positive impact on reading achievement; unsystematic variation among studies	+0.26**	[+0.22-0.31]
	Blok et al, 2002	CAI, AT	Phonological/ phonemic awareness	Ages 5-12.5	42	30 (71%)	1994	Small, positive effect, moderated by language of instruction and effect size at pretest	+0.19 (*k* = 50)	[+0.13-0.25]
	Kulik, 2003	CAI	Reading and math	Grades 2-8	9	3 (33%)	1994	Effects are lower for reading than mathematics	+0.06	*
	Jitendra et al., 2004	CAI	Vocab	Grades 4-12 with LD	5	5 (100%)	1988	Mixed effects; positive effects in four studies	+0.16 (*k*=2)	*
	Moran et al., 2008	CAI, ET, AT	Comp	Grades 6-8	20	16 (80%)	2002	Moderate, positive effect; higher effects for general education and comprehension focused interventions	+0.49	[+0.27-0.71]
	Strong et al., 2011	CAI	Word ID, Comp	Ages 6-11	6	5 (83%)	2007	No evidence that Fast ForWord is effective for improving vocab or overall reading	+ 0.08 -0.03	[-0.09-0.25] [-0.40-0.35]

Synthesis Type	Study	Tech Type	Outcomes	Participants	Studies Cited	Unique Studies	Median Year	Descriptive Findings	ES	Confidence Interval
Meta-analysis/ Best evidence synthesis	Slavin et al., 2011	CAI	Overall	K-5 struggling readers	12	1 (8%)	1994	CAI had few effects on reading	+0.09 (*k*=14)	*
	Cheung & Slavin, 2012a	ET, CAI	Overall	K-6 struggling readers	18	5 (28%)	2007	Small, positive overall effect favoring small-group integrated applications	+0.18 CAI (*k*=12) +0.06 FFW (*k*=2)	[+0.04-0.28] [-0.11-0.24]
	Cheung & Slavin, 2012b	ET, CAI	Overall	K-12	82	62 (76%)	2004	Small, positive overall effect	+0.11 CAI (*k*=56) +0.18 ITA (*k*=6)	[+0.07-0.15] [+0.08-0.28]

Note: Tech = Technology; CAI = computer-assisted instruction; ET = educational technology; AT = assistive technology; Overall = overall reading achievement; Comp = comprehension; Word ID = word identification; Vocab = vocabulary; LD = Learning Disabilities; FFW = Fast ForWord; ITA = Innovative Technology Applications * Not Available; **Converted from *r*

Table 2

Categorizations of Program Type

Synthesis Type	Study	Program Type (Technology)	Effect
Narrative	NRP, 2000	Replacement (AT)	Positive outcomes in both categories; hypertext
		Augmentation (CAI)	cited as a feature with promise
	Hall et al., 2000	Drill-and-practice (CAI)	No differences among program type
		Strategy instruction (CAI)	Positive effects of elaborate feedback
		Simulation (CAI)	
		Elaborate feedback (CAI)	
	Stetter & Hughes, 2010	Computer-presented vs. paper-presented (AT)	No difference with text presentation
		Hypertext support (AT)	Positive effects with supported text
Meta-analyses	Soe et al., 2000	Drill-and-practice (CAI)	Effects not disaggregated among program type
		Tutorial (CAI)	
		Dialogue (CAI)	
	Blok et al., 2002	Phonological awareness training (CAI)	No differences among program type
		Word reading with speech feedback (AT)	
		Flashed word reading (CAI)	
		Reading while listening (AT)	
		Mixed programs	
	Moran et al., 2008	Researcher-designed programs (ET/CAI)	+1.20 [CI: +0.91-1.49]
		Delivery systems (ET/AT)	+0.34 [CI:+ 0.19-0.49]
		Commercial programs (ET/CAI)	+0.28 [CI:+ 0.12-0.43]
	Cheung & Slavin, 2012a	Small-group integrated applications (ET)	+0.32 [CI: -0.05-0.69]
		Supplemental programs (CAI)	+0.18 [CI: +0.04-0.28]
		Comprehensive models (ET)	+0.04 [CI: -0.09-0.17]
		Fast ForWord (CAI)	+0.06 [CI: -0.11-0.24]
	Cheung & Slavin 2012b	Comprehensive (ET)	+0.28 [CI: +0.14-0.41]
		Computer-managed learning (ET/CAI)	+0.19 [CI: +0.02-0.36]
		Innovative technology applications (CAI)	+0.18 [CI: +0.08-0.28]
		Supplemental (CAI)	+0.11 [CI: +0.07-0.15]

Summary of results. Results of both narrative reviews and meta-analyses were all positive, with the exception of Strong et al. (2011). A median effect size of +0.13 indicates a small, positive overall effect. The median was consistent with the mean (*M* = 0.15; *SD* = 0.13; CI 0.07-0.24). Two potential outliers exist among this set of studies: Strong et al. (2011) found negative effects of CAI on reading, while Moran et al. (2008) found moderate effects (+0.49) of educational technology on the comprehension of middle school students. Most syntheses conducted after 2000 have utilized meta-analytic techniques to examine educational technology. Three recent meta-analyses of educational technology (Cheung & Slavin, 2012a; 2012b; Slavin et al., 2011) have disaggregated findings for CAI, finding slightly smaller effects for CAI than the median reported here in two studies, and slightly larger effects (+0.18) in their study investigating CAI with struggling learners. These studies have indicated that small-group sessions integrating technology produce the most benefit for all types of readers (range = +0.28-0.38; Cheung & Slavin 2012a; 2012b), though differing definitions of these models make it difficult to draw specific conclusions.

An examination of citations indicates a wide range of research represented among the reported syntheses. With almost 80% of cited studies reviewed in only one paper, a great deal of research from the past four decades has been analyzed over the last 15 years. The relative consistency in overall effect across such wide variation in studies provides preliminary evidence that CAI may have meaningful effects on the reading achievement of K-12 students. The narrative reviews and meta-analyses reported in this paper were, for the most part, methodologically sound; however, differing criteria for inclusion convolutes the aggregation of effects across studies, limiting the generalizability of overall findings.

Instructional Design Features of Included Syntheses

In addition to examining the overall effects of CAI in reading, a second question guiding this review involved investigating analyses of instructional design features shown to contribute to overall effects. Although none of the included syntheses reported the effects of specific instructional design features on reading outcomes, many studies categorized program type to disaggregate effects. Table 3 presents a summary of these categorizations. Among narrative reviews, supported text (i.e., hypertext with links to more information; NRP, 2000; Stetter & Hughes, 2010) and elaborate feedback (Hall et al., 2000) were cited as promising design features, based on studies reporting statistically significant differences for groups using CAI incorporating these features. Early meta-analyses categorized types of CAI in reading (Blok et al., 2002; Soe et al., 2000) but did not report differences among these program types. Moran et al. (2008) found large effects of researcher-designed programs and small but statistically significant effects of evidence-based delivery systems (e.g., integrating hypertext) and commercial programs. Cheung and Slavin (2012a; 2012b) found the smallest effects among supplemental CAI programs, but mixed results for technology integrating small- and large-group core instruction (i.e. comprehensive programs and small-group integrated applications). In sum, current syntheses have focused more on program type than on instructional design features that contribute to effects.

Discussion

This literature synthesis examined narrative reviews and meta-analyses to determine effects of CAI on reading outcomes and to analyze instructional design components that contribute to these effects. In this discussion, I provide a summary of overall results and discuss limitations of the current review. Gaps in current syntheses of research in reading CAI are discussed alongside the research questions leading to the meta-analysis reported in this paper.

Effectiveness of CAI

Synthesized results of four narrative reviews and nine meta-analyses indicate that CAI leads to positive outcomes in reading achievement. Effect sizes for CAI among the meta-analyses ranged from -0.03 to +0.49, with a mean effect of +0.15 (k = 12), which has been characterized as a small effect in educational research (Cohen, 1988). All narrative syntheses reported positive overall outcomes of CAI in reading. Previous syntheses of CAI effects on overall academic achievement have reached similar conclusions. For example, Kulik and Kulik's (1991) meta-analysis of 18 studies revealed a small, positive effect (+0.25), as did syntheses conducted by Becker (1992; ES = +0.18), and Ouyang (1993; ES = +0.18). The effect reported in Tamim et al.'s (2011) second-order analysis of existing meta-analyses was slightly larger (ES = +0.33), but consistent with previous syntheses, Tamim et al. (2011) included overall achievement outcomes rather than a specific content focus. The magnitude of the effects of CAI is smaller than those reported in meta-analyses of studies of more traditional methods of intervention delivery (e.g., one-to-one tutoring with a certified teacher),which have yielded moderate to strong effect sizes ranging from +0.41-+0.95 (Edmonds et al., 2009; Flynn, Zheng, Swanson, 2012; Scammacca et al., 2007; Swanson, 1999).

Though the small effect of CAI found in this synthesis might seem discouraging, the potential benefits of CAI, and the methodology used to examine those benefits, should be considered. For example, if the intention of a primary study is to provide evidence that CAI is as effective as traditional teacher-led instruction, comparing mean differences between treatment (CAI) and control (i.e., teacher-led instruction of the same material) may reveal lower estimates of effect (e.g., Wild, 2009). This type of comparison should be taken into account within meta-analyses, since even a small, positive effect would indicate that CAI could be used as an effective alternative to teacher-led instruction given personnel resource constraints. The differential effects of comparison type (i.e., CAI vs. no-treatment control or CAI vs. teacher-led instruction) were not reported in any of the included meta-analyses.

The positive mean effect of the nine meta-analyses, coupled with the positive results reported in the four narrative reviews, reveals overall improvements in reading outcomes following students' use of CAI. The consistency of this effect across decades of research provides further evidence that CAI can enhance reading outcomes of struggling learners, learners with identified disabilities, and learners on target for academic success.

Instructional Design Features

Instructional design features of CAI have been recommended through an examination of evidence-based practices in technology and reading literature (e.g., Moreno & Mayer 2007, Smith & Okolo, 2010). These recommended components, however, have either not been empirically analyzed in CAI synthesis research, or have been shown to have little to no effect on overall results. Empirical evaluation of program type revealed larger effects for researcher-designed programs (ES = +1.20; Moran et al., 2008) and applications integrated within small-group instruction (ES = +0.28-0.32; Cheung & Slavin, 2012a; 2012b). Mayer's work on instructional design features of CAI (e.g., Moreno & Mayer, 2007) offers empirical evidence of effective design features; however, their research has not specifically focused on reading.

In sum, empirical evidence of key instructional design features of CAI is limited; however, researchers have incorporated evidence gleaned from reading research with recommendations for technology. Examining CAI programs that contain instructional design components recommended in empirical reading literature allows educators to make evidence-based decisions when choosing technology, rather than making choices based on pricing or flashy interface designs.

Limitations and Implications for the Current Meta-Analysis

Though this literature review provides insight into the effectiveness of CAI in the area of reading across several syntheses, limitations of this literature review, the syntheses within, and the primary research represented helped guide my work in the current meta-analysis. One limitation of this literature review was the reliance on statistical calculations performed by other researchers. Reported effect sizes may be inaccurate or not fully representative of overall effects of CAI. Also, both Cohen's *d* and Hedge's *g* were reported and quantitatively synthesized. The nuances of these effect size calculations should be considered when interpreting the overall mean and median effect sizes reported (Cooper et al., 2009). Furthermore, effect sizes were not reported in the four narrative reviews; had these quantitative results been calculated, the mean and median effects reported in this paper might have been different.

Though search criteria and definitions of CAI varied among studies, a general research focus on the effects of instructional technology on reading achievement was characterized by all included studies. Based on this shared focus, it was surprising to find such wide variation in represented studies among the syntheses. The fact that 78% of primary studies were reviewed in only one synthesis indicates differences among study features (e.g., methodological criteria, definitions of CAI) that remain unexplained. For example, the stringent inclusion criteria of Slavin et al. (2011) and Cheung and Slavin (2012a; 2012b) eliminated studies based on factors such as short duration (<12 weeks), lack of two teachers per experimental group, and studies utilizing measures inherent to the intervention program (i.e., researcher-created assessments). Though their intent was to produce a best-evidence synthesis by including only high-quality designs (Slavin, 1986), a traditional meta-analysis allows researchers to analyze the effects of all variables, including methodological characteristics, publication type, and variation in program components (Cooper et al., 2009). The exclusion of primary studies based on research characteristics (e.g., poor methodological quality) may prevent a full understanding of the effects of CAI in reading. In an attempt to resolve this limitation, I have developed a concise methodology to include as many relevant studies as possible. A clear, operational definition of CAI will enhance the interpretations and generalizability of the findings of the current meta-analysis.

A significant gap in the research on CAI is the lack of empirical evidence from primary studies regarding the effectiveness of specific instructional design features (e.g., appropriate pacing, quality of feedback). Often, primary research articles do not provide descriptions of CAI in adequate detail for researchers attempting to code for design features, and the small sample sizes of many syntheses may not provide adequate power to detect effects. Future primary research should describe program features in enough detail that educators and researchers can analyze the instructional design components included. Component research of reading CAI, similar to that described by Moreno and Mayer (2007), should also be conducted in order to empirically establish effective design features. Moran et al. (2008) recommended developing a master codebook to serve as a heuristic for analyzing CAI. Until component research is conducted or future syntheses can determine program features that contribute to effects, a master codebook such as the one created by Bishop and Santoro (2006) might help educators sift through the vast array of flashy applications promising to enhance students' reading achievement. In this meta-analysis, I look within primary research to locate the recommended design components, summarize their presence in the literature, and analyze their contribution to overall effects and the ability of the instructional design components to explain variability among effects.

Another gap within CAI research is a lack of information regarding the areas of reading (e.g., phonics, vocabulary, comprehension) that may be taught more effectively through CAI. Though some syntheses focused solely on specific areas of reading (e.g., Blok et al., 2002; Moran et al., 2008), none of the other meta-analyses disaggregated effect sizes among different areas of reading. Most of the syntheses reporting results of comprehension outcomes discussed the use of assistive technology (e.g., text to speech devices) to improve overall comprehension (e.g., MacArthur et al., 2001; Moran et al., 2008). In fact, specific details on CAI to improve comprehension is lacking in this literature review. In this meta-analysis, I analyze the variability in effects based on specific reading outcomes. Given enough information from primary studies, moderator analyses could further inform this line of inquiry by determining not only areas of reading showing the most promise through CAI, but also the program features that contribute to specific reading outcomes. This type of information is critical for educators to determine what types of CAI are aligned with individual student need in order to target specific skill deficits.

A final limitation of the research on CAI in reading is somewhat difficult to remediate, as it relates to time. Technology has been changing rapidly across the decades, and continues to change every day. For example, the use of handheld devices, such as iPads, is increasing in schools today, and available syntheses have not examined research current enough to include information as to the effectiveness of these alternative devices. The years covered by syntheses in this review ranged from 1978 through 2011. The highest median year represented among syntheses was 2007, leaving approximately six years of current research to be analyzed. Furthermore, some of the primary studies analyzed among all but two meta-analyses were conducted over three decades ago. Blok et al. (2002) and Strong et al. (2011) were the only syntheses to include only research published after 1990. Further analysis of the cited publications revealed that 47.5% of cited studies were conducted during the 1990's, and less than 32% of cited studies have been conducted since 2000. The inclusion of studies that are 20+ years old must be considered, as the technology represented in overall effect size calculations is most likely very different from technology used in schools today. With a growing national emphasis on standards- and evidence-based practices, commercial CAI designers are creating programs that align with these practices, and may prove more effective than previously-used technology. In this meta-analysis, I will include current primary research (i.e., post-2000) and analyze the effects of publication year to determine changes within the last decade. Though no research on the effectiveness of handheld devices in reading was located for this meta-analysis, future syntheses should analyze these studies as they become available.

Implications and Research Questions

Though reading is a skill readily acquired by many, it is also a skill with which countless children and adults struggle. Providing students with reading interventions targeting specific skill deficits can allow for skill remediation, and an overall goal of gaining meaning from text can be attained (e.g., Flynn, Zheng, & Swanson, 2012; Scammacca et al., 2007; Swanson, 1999). Schools operating under an RTI framework are becoming well-versed in the need for interventions and the effects these interventions have on students; however, the provision of interventions can be difficult when personnel resources are constrained. Through CAI, students may receive individualized instruction targeting specific needs, increasing opportunities to respond, opportunities for feedback, and overall exposure to reading instruction. Furthermore, when a student works on targeted skills through CAI, licensed teachers may be freed up to work more closely with smaller groups of students on additional skills. The ability for flexible, small groupings in interventions incorporating CAI is an advantage of this method of intervention delivery.

When examining available interventions and weighing the costs and benefits of CAI versus traditional teacher-led intervention delivery, educators can look to the expansive research base and the accompanying research syntheses for evidence that CAI is a viable option for enhancing students' reading skills. In an era of increased focus on data- and evidence-based intervention delivery, CAI has the potential to supply a needed instructional resource providing targeted, individualized instruction to numerous students simultaneously. A comprehensive synthesis of primary research on CAI in reading conducted within the last decade is needed to examine potential changes in overall effect and to inform best practice on the instructional design features that contribute to those effects. Guided by the following six research questions, this dissertation seeks to examine not only current effects of CAI in reading, but also the overall quality of current research, thus informing the course of future research on this topic.

1. What is the mean effect of CAI on reading outcomes for students in preschool through high school and what is the evidence of variability among those effects?
2. What are the differential effects of CAI on reading outcomes across school levels (i.e., preschool/kindergarten, elementary, middle, and high school)?

3. What are the differential effects of CAI on reading outcomes across the five areas of reading (phonemic awareness, phonics, fluency, vocabulary, comprehension)?
4. What study characteristics contribute to the effects of CAI in reading (e.g., methodological characteristics, instructional components, research quality)?
5. What is the overall quality of the research according to the quality indicators proposed by Gersten et al. (2005)?

Chapter 3

METHOD

In an effort to locate all studies on the topic of CAI in reading for students in preschool through high school, I conducted a systematic search process followed by screening, coding, and analyses of effects. This chapter presents information on these procedures, including formulas used to compute relevant statistics. The chapter concludes with a description of the methods used to assess publication bias.

Study Retrieval

I conducted a systematic process of study retrieval to locate all relevant studies investigating CAI in reading. In an effort to capture a wide body of research on this topic, the search included both published and unpublished meta-analyses as well as non-peer-reviewed published research. The retrieval process was documented quantitatively within each step to provide further information about the constitution of research on this topic. An outline of the steps in the search procedure is provided in Appendix A.

Search procedure. An eight-step process of study retrieval was conducted to locate all relevant studies for this meta-analysis.

1. Systematic searches were conducted in four databases: Academic Search Premiere, Education Full Text, ERIC, and PsycINFO using the terms (*reading or literacy or language arts or phonics or phonemic awareness or phonological awareness or fluency or vocabulary or comprehension*) and (*instruction or instructional or intervention or education or RTI or response to intervention*). This subset of references was searched using the terms (*CAI*

or computer-assisted or computer-based or computerized or software or technology or tablet) and (*preschool or kindergarten or elementary school or middle school or high school or secondary or K-12 or grade or graders or students*). Citations and abstracts of the resulting articles were imported into Refworks, and duplicates were eliminated.

2. The titles and abstracts of all articles resulting from the first step of the search were screened according to broad criteria. Reports were eliminated if they did not include (a) an intervention study, (b) outcomes in reading, (c) CAI as defined in this dissertation, (d) a comparison group, or (d) CAI and/or assessments delivered in English (see Step 2b in Appendix A).
3. Remaining articles from the second step of the search were read in full, and articles were eliminated based on the inclusion criteria listed above (see Step 3b in Appendix A). In addition, articles reporting the same data in different formats (e.g., conference presentations and dissertations) were examined; the manuscript providing the most quantitative data was retained.
4. The reference lists of all included articles were examined for additional articles, and those were fully screened as in the third step of the search procedure.
5. Journals represented in the included articles were searched using the terms related to CAI, and resulting articles were screened in full.

6. IES reports available through the What Works Clearinghouse (http://ies.ed.gov/ncee/wwc) were examined to identify additional articles and technical reports, which were screened in full.
7. Google searches were conducted for each different type of CAI using the names provided for the programs in each report. Websites representing commercially available CAI were searched for additional technical reports, and resulting reports were screened for inclusion.
8. Researchers reporting results of noncommercial CAI (e.g., researcher-created) were contacted to locate additional unpublished studies. CAI was determined to be noncommercial if it was explicitly stated as such in the report or if the program could not be found online via a Google search. All additional studies were fully screened using the predefined inclusion criteria.

Eligibility criteria. Articles meeting the following six criteria were included in the sample of studies to be analyzed:

1. The study examined the use of CAI as defined in the introduction with a primary outcome measure of reading achievement, including specific measures (e.g., phonemic awareness) and/or broad reading measures. Studies that examined CAI including another subject (e.g., mathematics) without disaggregating effects of reading were excluded from this review.
2. The primary instruction was delivered through a computer (rather than teacher-led instruction); was interactive in that students provided responses through the keyboard, mouse, or touch screen; was responsive to the student

(i.e., program continuation dependent upon student interaction); provided instruction individually or in pairs (i.e., student to computer ratio no greater than 2:1); and was provided over some period of time (i.e., more than one session). Studies describing interventions using educational technology not meeting the definition of CAI (e.g., video-based instruction, interactive whiteboards) were excluded from this review, as were studies examining the use of assistive technology on reading outcomes.

3. The study was published or made available between 2000 and 2013 in order to capture the most current literature regarding CAI in reading.
4. The study participants included students in preschool through high school.
5. The study employed an experimental design with a comparison group and presented quantitative data that could be used to compute effect sizes.
6. The study used CAI provided in English in order to determine the differential effects of CAI on the five areas of reading determined by the National Reading Panel (2000).

Coding Procedure

Studies were coded within an Excel spreadsheet in the following categories: (a) report information, (b) sample characteristics, (c) participant demographics, (d) learner characteristics, (e) intervention information, (f) CAI information, (g) outcome characteristics, (h) results, and (i) quality indicators suggested by Gersten et al. (2005). The coding manual including variables within each category can be reviewed in Appendix B.

Report information. Variables related to study information were coded and analyzed. These variables included (a) the year the study was made available; (b) publication type (e.g., peer-reviewed journal, dissertation, technical report); (c) the country in which the study was conducted; (c) the funding source of the study (i.e., government or private funding); (d) the number of effect sizes included in the report; (e) the total number of participants in the report; and (f) the total number of participants in the treatment and control groups.

Sample characteristics. Coded variables related to the characteristics of the participants in the sample included (a) the total number of participants for the particular effect size being coded; (b) number of participants in the treatment and control groups; (c) the sampling method used; (d) the total attrition reported; and (e) reported attrition from the treatment and control groups.

Participant demographics. Coded variables related to the demographics of participants included (a) the school level of participants; (b) race; (c) the location of the participants and/or school in the study (i.e., urban, suburban, rural, or mixed; (d) the socio-economic status of the sample; and (e) gender. With regard to school level, the reported school grades associated with each effect were coded and a level was assigned. In the coding manual, Levels 1-4 represented students in (a) preschool through kindergarten (PreK-K); (b) elementary school (K-6), (c) secondary school (6^{th}-12^{th} grade); and (d) mixed levels (e.g., 1^{st}-12^{th} grades), respectively.

Results by school level are provided using the original levels assigned during coding; however, further analysis revealed an overlap in effects across school levels when studies included kindergarten participants. There was no overlap in effects between sixth grade students represented at the elementary (K-6) or secondary (6-8) levels. Furthermore, disaggregated effects were available for middle (6-8) and high school (9-12) participants. Thus, I conducted the moderator analyses with modified levels that represented students in (a) preschool through elementary school (Prek-6); (b) middle school (6-8); (c) high school (9-12); and (d) mixed (e.g., K-12).

Learner characteristics. Coded variables related to learning characteristics of the participants included (a) percentage of students reported to be English language learners (ELL); (b) percentage of students reported to be identified with specific disabilities; and (c) percentage of students reported to at-risk in the area of reading.

Intervention information. Coded variables related to the intervention used in the study included (a) how CAI was used within core literacy (i.e., integrated or supplemental); (b) nature of control condition (i.e., treated versus untreated controls); (c) length of sessions, sessions per week, and total duration of the intervention; and (d) type of teacher leading the intervention (i.e., licensed teacher, researchers or other school staff).

CAI characteristics. Coded variables related to CAI included (a) program name; (b) commercial availability of the software; (c) area(s) of reading addressed (e.g., phonics, vocabulary); (d) instructional design components (e.g., feedback, record keeping) and (e) structure of the CAI (i.e., linear or adaptive to student input). Though the structure of CAI is not included in design component suggestions (Bishop & Santoro, 2006; Moreno & Mayer, 2007), a review of primary research revealed a distinction between programs progressing in a linear fashion (e.g., students must master one section before continuing to the next section) or programs adapting to student input (e.g., changes in sequencing based on correct or incorrect responses).

Outcome measures. Coded variables related to outcome measures included (a) the name of the measure; (b) specific area(s) of reading measured (e.g., phonics, comprehension, broad reading); (b) type of measure (e.g., standardized, broad reading); and (c) whether or not the measure aligned with at least one component of reading reported for the CAI. The type of measure was coded as standardized if the report provided established technical characteristics demonstrating standardization of procedures. An outcome was coded as a broad measure if it represented more than one areas of reading within one measure. These variables were not mutually exclusive; thus, a measure could be coded as both a standardized and broad measure of reading.

Results. Coded variables related to results included (a) pretest means and standard deviations for treatment and control (when reported); (b) posttest means and standard deviations for treatment and control; (c) effect sizes and confidence intervals computed according to the formulas described below; (d) independence of the effects; (e) effect sizes reported in the study; and (f) reporting of scores (e.g., gain scores, mean posttest scores).

Quality indicators. The influence of methodological quality on study outcomes was examined by coding several variables, including the description of attrition within studies. Adapting the essential and desirable quality indicators described by Gersten et al. (2005), studies were coded on several variables of methodological quality, which are summarized in Table 4. Each variable, including design type (i.e., randomized) and attrition reporting, was given a score of 1 if present or 0 if absent. Scores were summed to determine an overall methodological quality score, with a maximum score of 12.

Table 4

Description of Quality Indicators (Gersten et al., 2005)

Category	Quality Indicator	Description
Participant Description	Abilities/disabilities described	Evidence to confirm the disabilities or difficulties presented
	Comparable groups	Evidence to confirm the comparability of groups across conditions
	Comparable interventionists	Evidence to confirm the comparability of interventionists across conditions
	Attrition	Attrition reported; less than 30%
Intervention Description	Intervention description	Intervention description clear and specific
	Fidelity of implementation	Fidelity assessed and described
	Control description	Control description clear and specific
Outcome Measures	Multiple measures	Multiple measures aligned with intervention and generalized
	Measures timed appropriately	Measures administered at appropriate time to capture intervention's effect
Data Analysis	Appropriate data analysis	Analysis linked to research questions and unit of analysis
	Effect size	Effect sizes reported and interpreted
Design	RCT	Randomized control trial

Gersten et al.'s (2005) recommendations regarding the application of quality indictors to special education research was intentionally vague, given the need for field tests of their recommended systems. Several publications (e.g., Baker, Chard, Ketterlin-Geller, Apichatabutra, & Doabler, 2009; Jitendra, Burgess, & Gajria, 2011; Jitendra et al., 2015) have applied the quality indicators in syntheses of research in writing, reading comprehension, and mathematics, respectively. Each of these publications focused on the quality of evidence of given practices, and used detailed rubrics to evaluate research reports. Though these detailed evaluations provide invaluable information regarding evidence-based practices in a particular domain, the investigation of research quality in this dissertation is not the sole focus of the meta-analysis; thus, the use of a dichotomous coding scheme (i.e., the indicator was met or not met) was deemed appropriate to provide a preliminary examination of research quality and the impact of quality on overall effects.

Interrater Agreement

Interrater agreement (IRA) was established for the screening and coding process following completion of the systematic search. I served as the first rater and a doctoral student and recent Ph.D. graduate of the Educational Psychology program at the University of Minnesota served as the second and third raters. During the screening process, 45% of the studies identified for full-text review were randomly selected and split between the two coders to establish IRA within the screening process. During the coding stage, all raters coded two studies together using the coding manual presented in a spreadsheet format (see Appendix B). Codes with low IRA were identified and refined. I coded all included studies in the meta-analysis, and the second and third raters independently coded a total of 25% of randomly selected studies. Studies analyzed by the additional coders did not overlap, thus, IOA was computed as described above, reported separately, and averaged across the studies selected for IOA. All disagreements between raters were discussed and resolved by consensus before analyzing coded data. IRA is reported in the results of this dissertation. Percentage of IRA was calculated by comparing coded data and applying the following formula:

$$\text{IRA} = \frac{\text{agreements}}{\text{agreements}+\text{disagreements}} \times 100$$

Excluded Studies

In order to enhance the replicability of this meta-analysis, as well as the inferences to be drawn from results, an analysis of excluded studies is provided in the final results through Table 6, which depicts studies excluded at each level of search and coding. Reasons for exclusion are reported alongside the number of studies excluded for each reason.

Analysis of Effects

In meta-analysis, the calculation of a common effect size metric is essential for standardizing effects and allowing for direct comparisons across studies. In the context of this meta-analysis, the effect size can also be thought of as the treatment effect, or the difference between treated and control groups; this measure provides an indication of the magnitude and direction (positive or negative) of intervention effects on participant outcomes (Borenstein, 2009; Lipsey & Wilson, 2001). Following calculation of effect sizes, further analyses can be conducted to draw inferences from the results. All calculations and analyses were conducted using Comprehensive Meta-Analysis (CMA), SPSS, and R Statistical Computing software.

Effect size extraction. Effect sizes were computed for all outcomes reported in the study. Several decision rules were applied when extracting effect sizes in order to preserve the assumption of independence of each effect (Cooper et al., 2009; Lipsey & Wilson, 2001). First, results of different outcome measures (e.g., phonemic awareness, comprehension) collected from the same sample of students were considered dependent. Thus, an effect size was computed for each measure and a single mean effect was included in the final analysis. When subtest scores were provided alongside an aggregated score, the aggregated scores were used to compute and report the effect size for that study, though the subtest score effects were also computed and coded to be used in moderator analyses. Second, independent samples contained within the same study (e.g., third grade and fifth grade) were treated as independent cases; effect sizes for all samples were included in the final analysis. Effects for dependent samples (e.g., effects of an at-risk subgroup of students) were computed; however only the aggregated mean effect size (e.g., all subgroups combined) was included in the final analysis. The computed effect sizes of dependent samples and/or outcomes were retained for moderator analyses within separate categories of variables (Cooper et al., 2009).

Effect size calculation. The standardized mean difference (Hedge's *g*) was used as the index of effect in this meta-analysis, and was computed using methods suggested by Borenstein (2009) and Lipsey and Wilson (2001). The effect sizes were calculated so that (a) a positive sign indicated the experimental group outperformed the control group, (b) a negative sign indicated the control group outperformed the experimental group, and (c) an effect of zero indicated no difference between the performances of either group. For primary studies providing relevant information, the standardized mean difference (d_j) of study *i* was calculated by subtracting the mean of the control group from the mean of the experimental group divided by the pooled standard deviation, such that:

$$d_i = \frac{\overline{Y}_{E_i} - \overline{Y}_{C_i}}{S_i} \qquad (1\text{-}1)$$

where:

$\overline{Y}_{E_i}$ = experimental group mean

$\overline{Y}_{C_i}$ = control group mean

S_i = pooled, within group standard deviation

The pooled, within group standard deviation was calculated using the following formula:

$$S_i = \sqrt{\frac{(n_{E_i}-1)S_{E_i}^2 + (n_{C_i}-1)S_{C_i}^2}{n_{E_i} + n_{C_i} - 2}} \quad (1\text{-}2)$$

where in the i^{th} study:

n_{E_i} = number of subjects in experimental group

n_{C_i} = number of subjects in control group

$S_{E_i}^2$ = variance of experimental group

$S_{C_i}^2$ = variance of the control group

In order to address the bias of *d*, particularly in overestimating effect sizes in small samples, Hedges' conversion (Hedges, 1981) was applied to estimate g_i. To convert from *d* to *g*, a correction factor, *J*, is used, such that:

$$J = 1 - \frac{3}{4df - 1} \quad (1\text{-}3)$$

where:

df = degrees of freedom used to estimate S_i or $(n_{E_i} + n_{C_i} - 2)$

Then,

$$g_i = J \: x \: d_i \quad (1\text{-}4)$$

where:

g_i = unbiased effect size estimate of d_i

When means and/or standard deviations were not provided, effect sizes were converted from other metrics using formulas recommended by Borenstein (2009).

Aggregation of effect sizes. In order to draw inferences about the results of any meta-analysis, researchers must choose between a fixed- or random-effects model of analysis. In a fixed-effects model, the distribution of effect sizes is homogeneous, thus dispersion around the mean is no greater than that expected from sampling error alone (Lipsey & Wilson, 2001). This model allows for inferences to be drawn about studies identical to the population of studies examined in the meta-analysis, with new research participants representing the only difference among studies (Shaddish & Haddock, 2009). In a random-effects model, inferences may be drawn about results of studies that have been conducted with new participants as well as changes in other study characteristics (e.g., setting, outcome measures; Shadish & Haddock, 2009).

Fixed-effects model. In a fixed-effects model, the mean effect is often computed by weighting each effect size (g_i) by the inverse of its variance (w_i) using a formula suggested by Lipsey and Wilson (2001). By applying this formula, each effect size is assigned a weight based on its sample size. Studies with larger sample sizes typically produce results with less sampling error (SE), thus providing more precise estimates of true effect (Lipsey & Wilson, 2001). The SE of each effect size (g_i) is calculated using the formula:

$$SE = \sqrt{\frac{n_{E_i} + n_{C_i}}{n_{E_i} * n_{C_i}} + \frac{g_i^2}{2\left(n_{E_i} + n_{C_i}\right)}} \qquad (1\text{-}5)$$

Then, each study effect size is weighted by the inverse of its variance, providing a direct estimate of the precision of each effect size through the SE:

$$w_i = \frac{1}{SE_i^2} \quad (1\text{-}6)$$

where:

w_i = calculated weight of study j

SE_i = standard error of g for study i

The estimation of the weighted mean effect ($\bar{g}$), is the sum of each individual effect size weighted by the inverse of its variance, such that:

$$\bar{g} = \frac{\sum(w_i * g_i)}{\sum w_i} \quad (1\text{-}7)$$

Random-effects model. Given the variability in study characteristics among studies of CAI in reading (e.g., differing interventions, grade levels, settings), I hypothesized that each observed effect size differs from the population mean by both sampling error and a second variance component associated with random effects. In the random-effects model, total variance associated with the distribution of effect sizes is the sum of these two variance components, such that:

$$v_i^* = v_\theta + v_i \quad (1\text{-}8)$$

where v_θ is the estimate of between-studies variance and v_i represents the estimate of within-study variance due to sampling error (Lipsey & Wilson, 2001). Under a random-effects model, the weighted mean (equation 1-7) is computed including this unconditional variance estimate with formulas suggested by Cooper et al. (2009), and confidence intervals are recalculated. Application of a random-effects model and the associated calculations were conducted via CMA software.

Quantifying heterogeneity among effects. A Q test of homogeneity was conducted to determine the homogeneity of the distribution of effect sizes. A rejection of the null hypothesis indicates that the variability of effect sizes is larger than would be expected from sampling error alone. A statistically significant Q statistic, therefore, indicates a heterogeneous distribution, challenging the assumption of a fixed effects model, and providing further evidence supporting a random-effects model. The Q test was performed using formulas suggested by Lipsey and Wilson (2001) and Cooper et al. (2009) through CMA software.

While the Q test can confirm whether or not a set of effects are homogenous, the I^2 index can help quantify the variance among effects. The I^2 index represents the percentage of total variation in a set of effect sizes due to true heterogeneity or between-studies variance, rather than chance (Huedo-Medina, Sanchez-Meca, Marin-Martinez, & Botella, 2006). I^2 can be thought of as a measure of inconsistency across studies' findings (Borenstein, Hedges, Higgins, and Rothstein; 2009; Higgins, Thompson, Deeks, & Altman; 2003). I^2 was computed using CMA software with formulas suggested by Higgins & Thompson (2002) and Borenstein, et al. (2009) such that:

$$I^2 = \left(\frac{Q - df}{Q}\right) \times 100\% \tag{1-9}$$

Negative values of I^2 are truncated to zero when the degrees of freedom are greater than the Q value. I^2 lies between 0% and 100% with 0% indicating no observed heterogeneity. Higgins et al. (2003) suggest that an I^2 value of 25% low indicates low observed heterogeneity, 50% moderate, and 75% substantial evidence of heterogeneity (Borenstein et al., 2009).

Calculating confidence intervals. The precision of each mean effect estimate, or the range within which the population mean is likely to be (Lipsey & Wilson, 2001) was determined by computing the estimated SE of the mean as the square root of the sum of the inverse variance weights to calculate a 95% confidence interval (α = .05), such that:

$$SE_{\bar{g}} = \sqrt{\frac{1}{\sum w}} \quad (1\text{-}10)$$

where:

Lower limit of 95% *CI* = $\bar{g} - 1.96 * SE$

Upper limit of 95% *CI* = $\bar{g} + 1.96 * SE$

Confidence intervals resulting from random-effects models were computed using CMA software and reported within the final results.

Moderator analyses. Using CMA and R Statistical Computing software, the contribution of categorical and continuous variables on overall effects was examined. A categorical moderator analysis, analogous to an analysis of variance, can highlight variables that systematically explain variation in effects. Categorical variables (e.g., school level, instructional design components) were examined to determine the ability of particular variables to explain variation in the distribution of effect sizes. The effect sizes represented within categories of school level and areas of reading were disaggregated by category and examined separately to determine the magnitude of effects within these subsets of data. CMA software was used for all categorical moderator analyses, using a random-effects model.

Continuous (e.g., duration of treatment) and categorical variables were also analyzed using meta-regression to identify models explaining variation in effects (Lipsey & Wilson, 2001). Meta-regression was conducted using the Metafor package in R (Viechtbauer, 2010a) using restricted maximum-likelihood estimation. Procedures suggested by Viechtbauer (2010b) were used to guide these analyses and the interpretation of results.

Addressing extreme effects. If the distribution of effect sizes indicates extreme effect size values that are noticeably discrepant from most other effects, these extreme effects may have disproportionate influence on the statistics computed in the analysis (Lipsey & Wilson, 2001). Analysis of the distribution of effects with and without these outliers allows for comparison of the trimmed versus untrimmed distribution, lending to an overall interpretation and justification for the inclusion or removal of extreme effects (Hunter & Schmidt, 2004). Effect sizes larger than ± 2.00 standard deviations from the mean effect size were recoded for accuracy. Elimination of outliers that are extreme simply because of large sampling errors can result in overcorrection for sampling error and underestimation of the standard deviation (Hunter & Schmidt, 2004). Considering these implications, only the most extreme effect sizes ($\pm 3.00\ SD$) were considered for removal, and analyses including and removing these outliers were compared, providing an indication of how sensitive the combined measures of effect were to extreme effect sizes (Greenhouse & Iyengar, 2009).

Publication Bias

Publication bias represents a major threat to the validity of the results of a meta-analysis, since published studies tend to be selected due to the presence of significant outcomes (Sutton, 2009). The inclusion of published and unpublished studies (e.g., journal articles, technical reports, dissertations) can aid in the reduction of publication bias in this study; however, three methods of testing for publication bias strengthen this evidence. First, mean effects from published (or peer-reviewed) vs. unpublished studies were compared. Second a funnel plot was generated to examine symmetry of effects around the underlying true effect. If publication bias was present, a gap might exist in one corner of the funnel indicating suppression of smaller, unfavorable, or nonsignificant studies (Sutton, 2009). Third, the classic fail-safe N was computed among the independent effect sizes in this analysis to assess for publication bias (Sutton, 2009). The fail-safe N of *x* means an additional *x* number of studies would need to be included in order to nullify the effect found in this analysis. Once *x* is computed, a conclusion can be reached regarding whether it is reasonable to assume that *x* number of unpublished studies exist in the literature.

Chapter 4

RESULTS

The purpose of this study was to examine the effects of CAI on the reading outcomes of students in preschool through high school. To answer the proposed research questions, summary effect sizes were computed and examined across school levels (preschool, elementary, middle, and high school) and by outcomes in reading (phonemic/phonological awareness, phonics, fluency, vocabulary, and comprehension). The contribution of moderators (e.g., methodological characteristics) to the variation in effect sizes was examined using categorical moderator analysis and meta-regression. The overall quality of research was also examined and analyzed as a potential moderator of effect size magnitude.

The effect sizes reported in this study represent three unique sets of data including comparisons of CAI with control groups receiving (a) no treatment or non-reading CAI reporting posttest scores (Dataset A); (b) no treatment reporting gain scores (Dataset B); and (c) traditional, teacher-led intervention reporting posttest scores (Dataset C). Comparisons of CAI with teacher-led intervention reporting gain scores were not located for this review. Quantitative analyses of these datasets were conducted separately.

In this chapter, I begin with an overview of the study retrieval process, inter-rater agreement among coders, and descriptive information of the included reports. Then, I examine extreme effects and decisions to include or exclude potential outliers. I describe the results of the analyses according to each research question. Next, I describe the quality of research represented in the included reports and examine the impact that research quality had on overall effects. I conclude with the results of an analysis to assess publication bias in the results.

Study Retrieval

The systematic search of Academic Search Premiere, Education Full Text, ERIC, and PsycINFO yielded a total of 2,189 reports. Table 5 provides a summary of studies excluded and retained at each stage of the retrieval process. Following the database search and the full-text screening, 33 reports were retained. The reference lists of these reports were examined, as were the reference lists of the research syntheses described in Chapter 2 (see Table 2, p. 40). This stage of the retrieval process yielded an additional 25 reports. The journal search and WWC report search (http://ies.ed.gov/ncee/wwc) yielded an additional 10 reports. The website search identified 2 reports, though none were retained, and researcher contact yielded no additional reports due to a lack of response from the contacted authors. Following coding, 61 reports were retained for inclusion in the meta-analysis.

Table 5

Number of Reports Included/Excluded at Each Stage of the Retrieval Process

Stage	Reports Identified	Reports Excluded	Reports Retained	Total Reports Progressing to the Next Stage
Database Search	2,189	689	1,500	1,500
Abstract Screening	1,500	1433	67	67
Full-Text Screening	67	34	33	33
Reference List Review	40	15	25	58
Journal Search	5	2	3	61
WWC Report Search	8	1	7	68
Website Search	2	2	0	68
Contact Researchers	0	0	0	68
Coding	68	7	61	61

Table 6 provides a summary of the reasons for report exclusion. Of the reports excluded, 974 reports (45%) were not intervention studies and 713 reports (33%) were duplicates between databases. The remaining reports were excluded because they did not report outcomes in reading (10%) or did not examine CAI as defined in this study (5%). Less than 3% of the identified reports did not have an adequate control group, were not in English, did not have enough data to compute effect sizes, or examined students at the college level. A total of 2,183 reports were excluded during the retrieval process.

Table 6

Number of Excluded Reports with Reason for Exclusion (N = 2,183)

Exclusion Reason	Number of Studies	Percent of Total
Not an intervention study	974	45%
Duplicates from databases	713	33%
Not focused on reading	225	10%
Not CAI as defined	117	5%
Inadequate control group	72	3%
Not in English	42	2%
Not enough data	32	1%
Not PreK-12	7	<1%
Total Studies Excluded	*2,183*	*100%*

Interrater Agreement

I screened, reviewed, and coded all studies included in the retrieval process and the final meta-analysis. During screening and coding, a graduate student and recent graduate of the Educational Psychology doctoral program at the University of Minnesota served as additional coders to establish inter-rater agreement. Of the studies identified for full-text review as part of the screening process, 30 studies (45%) were randomly selected to be screened by one of the two coders. Inter-rater agreement was 87% with each respective coder during screening. Inter-rater agreement was also assessed for 15 (25%)

of the 61 coded articles. Inter-rater agreement averaged 85% with the first coder and 88% with the second coder. Inter-rater agreement among the quality indicators averaged 83% with the first coder and 85% with the second coder. All disagreements were reviewed, resolved, and modifications to coded variables were made when appropriate.

Descriptive Summary of Included Reports

Table C1 in Appendix C provides a breakdown of reports by CAI name (listed alphabetically) for participants in preschool through fifth grade. Table C2 provides these characteristics for participants in sixth through twelfth grade. Table D1 in Appendix D provides a summary of CAI characteristics, including areas of reading addressed, instructional components included, the structure of the technology (e.g., linear or adaptive) and dependent variables assessed. These tables are intended to provide further information regarding individual reports and instructional design components included in each CAI program reported in this set of studies. In the remainder of this chapter, I will synthesize effects across studies, rather than focusing on individual reports.

Report characteristics. Following study retrieval and coding, a total of 61 reports yielding 532 effect sizes were included in this meta-analysis. Table 7 presents a descriptive analysis of report characteristics. Publication year ranged from 2000 to 2013 (*Mdn* = 2007) with at least one report published in each year. Journal articles represented the most common publication type (69% of all included reports). Though several conference papers were located during the study retrieval process, these reports were found to be published elsewhere and were thus excluded as duplicates. No master's theses or unpublished documents were located for this review. The vast majority of included studies were conducted in the United States (87%). Though most reports were conducted in mixed or unspecified regions (30%), a similar percentage (26%) of reports were located in urban areas. Over half of the studies were conducted exclusively within the elementary level (K-5), though many studies contained participants in mixed school levels (20%).

Table 7

Descriptive Characteristics of Included Reports (N = 61)

	Descriptors	*n*	Percent of Total
Year of Publication	2000-2004	20	33%
	2005-2009	22	37%
	2010-2013	19	31%
Publication Type	Journal Articles	42	69%
	Dissertations	12	20%
	Technical Reports	7	11%
Country of Study	USA	54	89%
	Great Britain/England	6	10%
	China	1	<1%
	Finland	1	<1%
Location of Study	Urban	16	26%
	Suburban	9	15%
	Rural	6	10%
	Mixed/Other/Not Specified	30	49%
Funding Source	Government	18	29%
	Private	13	21%
	Other/Not Specified	31	50%
School Level	Preschool-Kindergarten	6	10%
	Elementary (K-6)	32	52%
	Secondary (6-12)	12	20%
	Mixed Levels	12	20%

Note: Overlap in school level resulted in some reports included in multiple school levels

Sample characteristics. Table 8 presents a summary of participant sample characteristics. A total of 38,940 participants were included across 61 reports. Total sample size ranged from 20 to 16,143 (*M* = 638.4, *SD* = 2,190). Twelve reports included less than 50 participants, 18 reports included 50-100 participants, 22 reports included 100-500 participants, and 11 reports included more than 500 participants.

The number of participants making up the treatment groups (n = 17,255) was less than the number of participants in the control groups (n = 21,685). On average, studies reported samples consisting of 37% White students, 29% African American students, 30% Hispanic students, and 9% Asian/Pacific Islander students. Boys tended to outnumber girls by an average of 12%. When reported, an average of 60% of the samples represented students from low socio-economic backgrounds as measured by free-and-reduced lunch, 84% were considered at-risk in reading, and an average of 27% of samples had identified learning disabilities.

Table 8

Characteristics of Participant Samples Across Reports (N = 38,940)

	Descriptors	% of Sample
Sample	Treatment	44%
	Control	56%
Race[a]	White	37%
	African American	29%
	Hispanic	30%
	Asian/Pacific Islander	8%
	Other/Not Specified	9%
Gender[a]	Boys	56%
	Girls	44%
Status[a]	Low SES	60%
	At-Risk	84%
	ELL	34%
	Learning Disabilities	27%

Note: SES = socio-economic status as determined by free-or-reduced lunch
[a]Reported percentages were averaged across studies

Intervention characteristics. A summary of intervention characteristics is presented in Table 9. These characteristics are drawn from the independent effects (k =

101) extracted from each data set; thus, there is some overlap among studies that reported results for multiple comparison groups. The contribution of these characteristics on overall effects will be discussed further in the results of the meta-analysis.

Of the extracted effects, the vast majority compared CAI to untreated controls or non-reading CAI (82%). Comparisons of CAI to teacher-led intervention in reading comprised 13% of the effects. Interventions were most often conducted in a separate room (47%), though 21% reported CAI use within the core literacy classroom. Almost a quarter of extracted effects did not contain information regarding the location of the intervention.

The nature of CAI use was distributed relatively evenly between CAI integrated into core literacy (i.e., no additional reading time), supplemental to core literacy (i.e., additional reading time), and mixed or not specified. The majority of sessions lasted up to 30 min (62%) and 16% of studies reported session lengths extending beyond one hour. The maximum session length was 180 min (M = 40.8, SD = 34.3). Most interventions occurred three to five days per week (54%), though almost one-quarter of extracted effects did not contain information regarding the number of sessions per week. The total duration ranged from 2-88 hrs (M = 24.1, SD = 21.4), with duration split relatively evenly among interventions lasting 1-10 hrs, 11-20 hrs, and 21-50 hrs. Information on total duration was not available for 16% of the extracted effects.

Table 9

Intervention Characteristics across Data Sets (k = 101)

Characteristics	Descriptors	*k*	% of Total
Control Type	No Treatment	76	75%
	Non-Reading CAI	6	6%
	Traditional Intervention	13	13%
	Other/Not Specified	6	6%
Location of Intervention	Core Literacy Classroom	21	21%
	Separate Room	47	46%
	Mixed	8	8%
	Home	2	2%
	Other/Not Specified	23	24%
Literacy Component	Integrated	34	33%
	Supplemental	31	30%
	Mixed/Other/Not Specified	36	35%
Session Length	0-30 min	63	62%
	30-60 min	11	11%
	60+ min	16	16%
	Not Specified	11	11%
Sessions Per Week	1-2	21	21%
	3-5	55	54%
	Not Specified	23	23%
Total Duration	0-10 hrs	26	26%
	11-20 hrs	29	29%
	21-50 hrs	21	21%
	51-100 hrs	9	9%
	Not Specified	16	16%
Teacher	Licensed Teacher	49	49%
	Other School Staff	8	8%
	Researcher	18	18%
	Not Specified	26	26%
CAI	Commercially Available	89	88%
	Not Commercially Available	12	12%
Areas of Reading Addressed	Phonemic/Phonological Awareness	63	62%
	Phonics	56	55%
	Fluency	43	43%
	Vocabulary	45	45%
	Comprehension	57	56%
Instructional Design Components	Guided Activity	20	20%
	Corrective Feedback	73	72%
	Elaborate Feedback	12	12%
	Pacing	59	58%
	Pretraining	5	5%
	Record Keeping	75	74%
CAI Structure	Linear	65	64%
	Adaptive	22	22%
	Not Specified	14	14%

All areas of reading were addressed relatively evenly by the CAI programs, though phonemic/phonological awareness was represented most often (62%). Over 72% of extracted effects represented CAI addressing two or more areas of reading and 17% of effects represented CAI addressing all five areas of reading. All coded intervention components were represented among the extracted effects, with record-keeping (74%), corrective feedback (72%), and student-controlled pacing (58%) represented most often. Over 71% of effects represented two or more instructional design components, with 22% representing four or more components. Only one study (Hasselbring & Goin, 2004) addressed all six coded components.

Characteristics of dependent variables. Table 10 summarizes the areas of reading assessed among the extracted effects. Phonemic/phonological awareness, phonics, and comprehension were each assessed in approximately 50% of effects. Fluency and vocabulary were represented relatively less often. Over 11% of extracted effects did not have information about the areas of reading assessed. For those effects with dependent variable information, 58% reported outcomes in multiple areas, and 32% reported outcomes in one area of reading. No extracted effects represented a combination of all five areas of reading. Almost all effects included standardized measures (i.e., measures using a standard protocol and scoring procedure), and half of all effects included a broad measure of reading (i.e., measuring one or more areas of reading with one assessment).

Table 10
Descriptive Characteristics of Dependent Variables (k = 101)

	Descriptors	*n*	% of Total
Area of Reading	Phonemic/Phonological Awareness	47	47%
	Phonics	47	47%
	Fluency	13	13%
	Vocabulary	25	25%
	Comprehension	50	50%
Numbers of Areas Assessed	Information not provided	11	11%
	1	32	32%
	2	30	30%
	3	22	22%
	4	6	6%
	5	0	0%
Type of Measure	Standardized	96	95%
	Broad Measure of Reading	50	50%

Extracted Effect Sizes

Once the 533 effect sizes were extracted from 61 reports, they were split into three sets of data, as represented in Table 11. Several reports provided enough information to compute effect sizes for more than one dataset; analyses conducted within each set of data maintained the independence of those effects. Reports containing comparisons with no treatment or non-reading CAI and reporting posttest scores (Dataset A) represented the largest set of effects in this meta-analysis. Rather than representing their own dataset, the five reports comparing treatment to controls receiving non-reading CAI were included in Dataset A, since comparisons of outcomes with non-reading CAI and business-as-usual control groups represented the effects of CAI in reading versus no CAI in reading. There were no non-reading CAI control groups in Dataset B.

Reports containing no-treatment comparisons and reporting gain scores (Dataset B) were less common and resulted in a much smaller dataset of independent effects. One report in Dataset B (Rasinski, Samuels, Hiebert, Petscher, & Feller, 2011) contributed 160 total effect sizes and seven independent effects. Reports containing comparisons of CAI with teacher-led intervention and reporting posttest scores (Dataset C) also resulted in a small set of independent effect sizes. There were no reports comparing CAI to teacher-led intervention using gain scores as the measure of effect.

Table 11

Description of Datasets used in Meta-Analysis

Dataset Reference	Comparison	Score Type	Reports (*n*)	Total Effect Sizes (*k*)	Independent Effect Sizes (*k*)
A	No Treatment/ Non-Reading CAI	Posttest	49	263	66
B	No Treatment	Gain	8	217	22
C	Teacher-led Intervention	Posttest	11	52	13

Several reports contained multiple outcomes for independent groups of students. In these cases, multiple outcomes were combined according to procedures suggested by (Borenstein et al., 2009) to create a mean effect for each independent subgroup. Two reports separated posttest data by subgroups (e.g., students with and without disabilities). In these cases, the effects of the subgroups were combined according to procedures suggested by (Borenstein et al., 2009) to create one mean effect across subgroups. In all above cases, all coded effects (i.e., independent and non-independent) were retained for moderator analyses within each dataset.

Five studies included several treatments (i.e., Fast ForWord and Earobics) compared to one control (see Tables C2-C3, Appendix C). Though these comparisons were not independent of one another, analyses were conducted separately for each comparison and with all included comparisons. The inclusion or exclusion of these non-independent subgroups had very little impact on the mean effect, variance, or confidence intervals of the analyses; thus, all comparisons in these four reports were treated as independent comparisons.

Addressing Extreme Effects

Outlier analyses were conducted by first plotting the distribution of effects (see Figs. 1-3) among each of the three data sets. The standard deviation (*SD*) of each effect from the mean effect was then analyzed, and effect sizes $\pm$2.00 *SD*s from the mean effect were recoded for accuracy. The most extreme effects ($\pm$3.00 *SD*s from the mean) were considered for removal, and analyses including and removing these outliers were compared.

Dataset A. The distribution of effect sizes in Dataset A is provided in Figure 1. The distribution appeared normal with one potential outlier. Further analysis revealed two effects sizes greater than 2 *SDs* from the mean. Gale's (2006) subgroup of kindergartners contributed an effect size of 1.60, which was 2.35 *SD*s from the mean. This effect size was recoded and found to be accurate. Hansen, Llosa, and Slayton's (2004) subgroup of first graders using Waterford Early Reading Program (WERP) contributed an effect size of -2.16, which was 4.36 *SD*s from the mean. This effect size was recoded and found to be accurate. Analyses including and removing Hansen et al.'s (2004) extreme effect were conducted and resulted in an effect size difference of 0.02. The confidence intervals of both analyses remained statistically significant, with relatively little change in the width of the confidence interval. Given the relative quality of this particular study (i.e., 75% of indicators met), the large sample size of the subgroup (n = 1,015) and the statistical significance of the effect [95% CI: -2.34, -1.98], this effect was included in all further analyses.

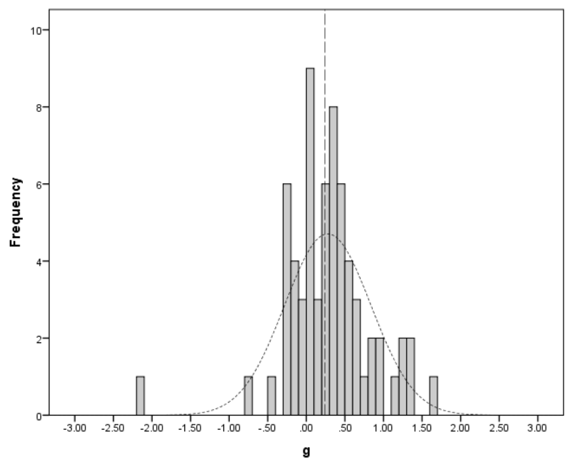

Figure 1. Frequency distribution of Dataset A with normal distribution and weighted mean as reference lines (M = 0.24; k = 66).

Dataset B. The distribution of Dataset B can be examined in Figure 2. Though the distribution is not normal, with so few effects (k = 22) it is difficult to discern the shape of the distribution. Further analysis revealed two effect sizes from one report greater than 2 *SD*s from the mean. Greenlee's (2001) study of Academy of Reading included three independent subgroups of students. The effect size obtained by the second-grade students was -0.94 (-2.57 *SD*s from mean) and the effect size obtained by the third-grade students was +0.90 (2.29 *SD*s from mean). These effects were recoded and found to be accurate and were thus included in all further analyses.

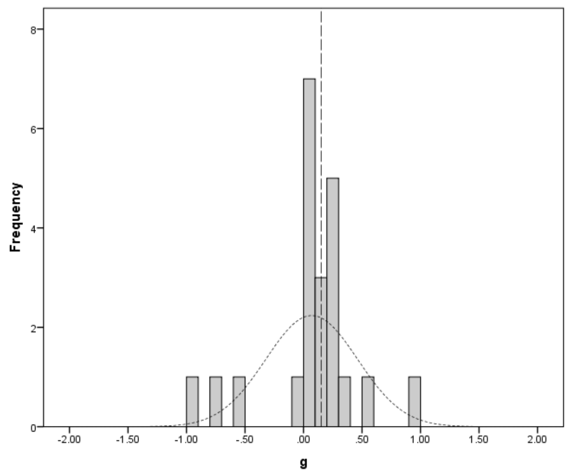

Figure 2. Frequency distribution of Dataset B with normal distribution and weighted mean as reference lines (*M* =0.15; *k* = 22).

Dataset C. The distribution of Dataset C can be examined in Figure 3. The distribution has a positive skew with one potential outlier. Further analysis revealed one effect size greater than two *SD*s from the mean. Cole and Hilliard's (2006) examination of Reading Upgrade revealed an effect size of +2.29 (2.92 *SD*s from mean). This effect size was recoded and found to be accurate. Since this effect was approximately three *SD*s greater than the mean effect for this dataset, analyses including and removing the effect were conducted. Among independent effects, inclusion of this study resulted in an effect size difference of 0.13 favoring inclusion of the report with neither mean effect reaching statistical significance. Analysis of all effects revealed negative mean effects when including or excluding the Cole and Hilliard (2006) study. The distribution of effects removing this outlier remained somewhat positively skewed. Given the small sample of effects in Dataset C and the relatively minor difference between mean effects including and excluding Cole and Hilliard (2006), further analyses included this report. The characteristics of Cole and Hilliard's (2006) report are included in Table C1 in Appendix C.

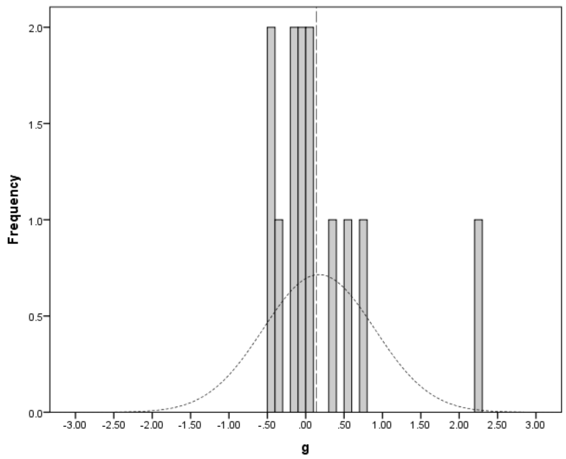

Figure 3. Frequency distribution of Dataset C with normal distribution and weighted mean as reference lines (M = 0.14; k = 13).

Results of the Meta-Analyses

During the coding process, 61 reports contributed 532 effect sizes. Of these effects, 101 independent effects were extracted and divided into datasets according to the source of the effect (see Table 11, p. 82). Results were analyzed using a random effects model, as described in Chapter 3. Results of tests of heterogeneity (Q and I^2) are provided for each research question. The Q-value provides a test of significance of the heterogeneity of the effect sizes while I^2 quantifies the percentage of total variability in a set of effect sizes due to between-studies variability (Huedo-Medina et al., 2006). Higgins et al. (2003) suggest that I^2 values of 25%, 50%, and 75% can be considered as low, moderate, and high respectively. This section is structured according to each of the five research questions guiding this meta-analysis. In some places, figures are provided to more clearly illustrate relations between variables.

What is the mean effect of CAI on reading outcomes for students in preschool through high school and what is the evidence of variability among those effects? To answer this research question, random-effects analyses were conducted with all uncombined effects (i.e., all subgroups, comparisons, and outcomes treated as independent) and with independent effects from each report. Table 12 provides a statistical summary of the effects from each of the three datasets. The forest plots in Figures 4-6 provide a visual representation of each independent effect size and its confidence interval and allows for comparisons between individual point estimates and the overall summary effect size.

Table 12

Statistical Summary of Effects of CAI on Reading Outcomes

Descriptor	*k*	*g*	SE	Var	95% CI	*Q*-Value	I^2
Dataset A							
All Effects	263	+0.21	0.06	0.004	[0.09, 0.33]	9275.22*	97%
Independent Effects	66	+0.24	0.07	0.005	[0.11, 0.38]	1080.88*	94%
Dataset B							
All Effects	217	+0.13	0.02	0.000	[0.10, 0.16]	870.63*	75%
Independent Effects	22	+0.15	0.04	0.001	[0.08, 0.22]	48.70*	57%
Dataset C							
All Effects	52	-0.05	0.06	0.003	[-0.16, 0.07]	159.23*	68%
Independent Effects	13	+0.14	0.15	0.021	[-0.15, 0.42]	52.43*	77%

*$p<0.005$

Dataset A. This dataset was made up of studies examining CAI compared to no-treatment controls or controls receiving non-reading CAI. Posttest scores were used to compute effect sizes. Dataset A consisted of 49 reports contributing 263 total effects and 66 independent effects. The model including all effects (k = 263; g = +0.21) and independent effects only (*k* = 66; *g* = +0.24) were similar in magnitude and both statistically significant (see Table 12). The statistically significant *Q*-values and I^2 ranging from 94%-97% indicate significant heterogeneity between effects. The forest plot in Figure 4 also demonstrates this heterogeneity, with many non-overlapping confidence intervals among individual effects.

Dataset B. This dataset was made up of studies examining CAI compared to no treatment controls with gain scores used to compute effect sizes. Dataset B consisted of eight reports contributing 217 total effects and 22 independent effects. The model including all effects (*k* = 217; *g* = +0.13) and independent effects only (*k* = 22; *g* = +0.15) were similar in magnitude and both statistically significant (see Table 12). The statistically significant *Q*-values and I^2 ranging from 57%-75% indicate moderate to substantial heterogeneity among these effects. The forest plot in Figure 5 shows this varied overlap in confidence intervals among individual effects in Dataset B.

Dataset C. This dataset was made up of studies examining CAI compared to control groups receiving teacher-led interventions in reading. Posttest scores were used to compute effect sizes. Dataset C consisted of 11 reports contributing 52 total effects and 13 independent effects. The model including all effects (*k* = 52; *g* = -0.05) and independent effects only (*k* = 13; *g* = +0.14) were different in magnitude and direction, and neither were statistically significant (see Table 12). After removing the large effect size extracted from Cole and Hilliard (2002), the mean of the independent effects dropped to +0.01. The statistically significant *Q*-values and I^2 ranging from 68%-77% indicate moderate to substantial heterogeneity among these effects. The forest plot in

Figure 6 demonstrates overlap in the confidence intervals of the individual point estimates.

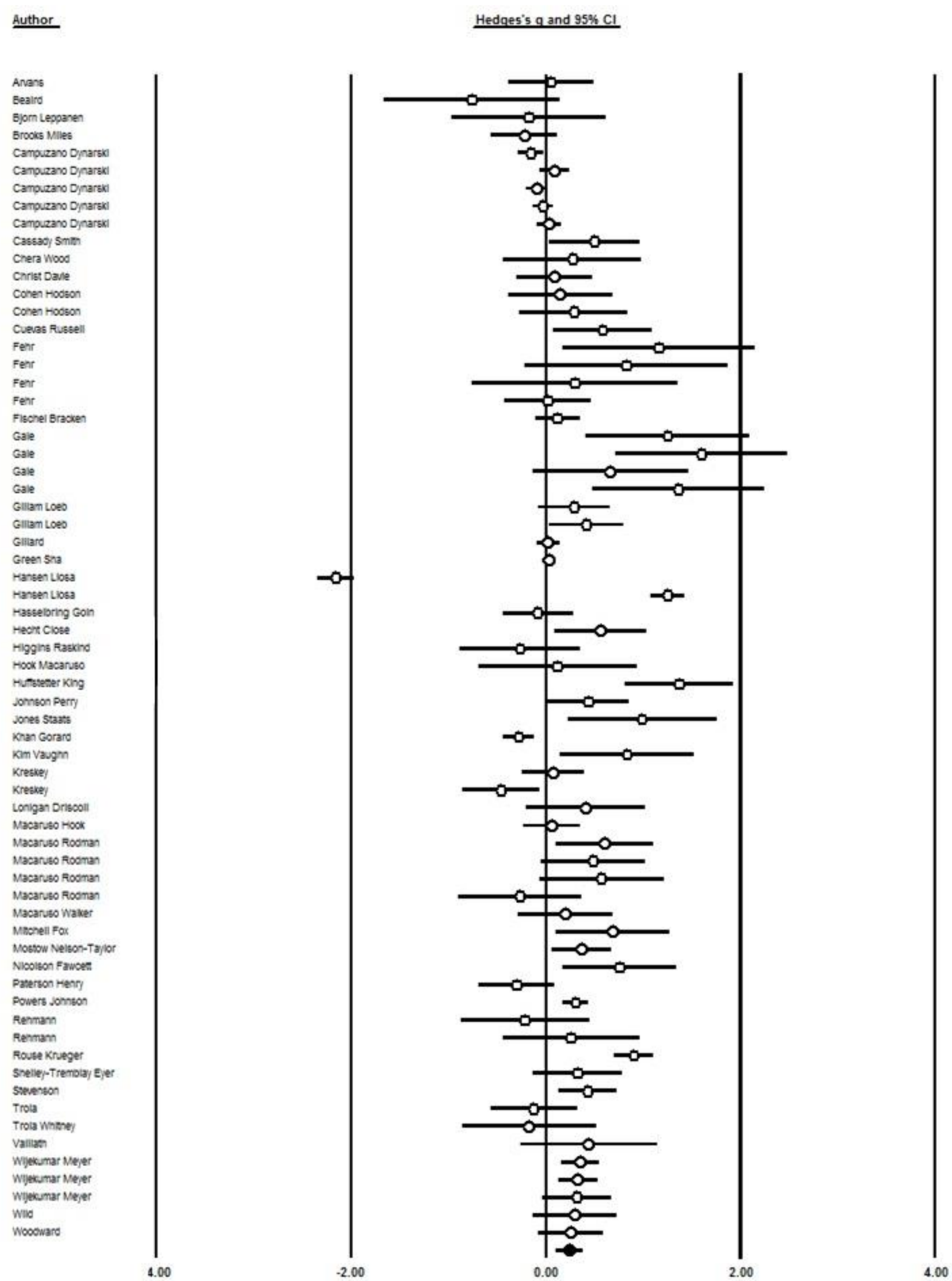

Figure 4. Forest plot of independent effects in Dataset A

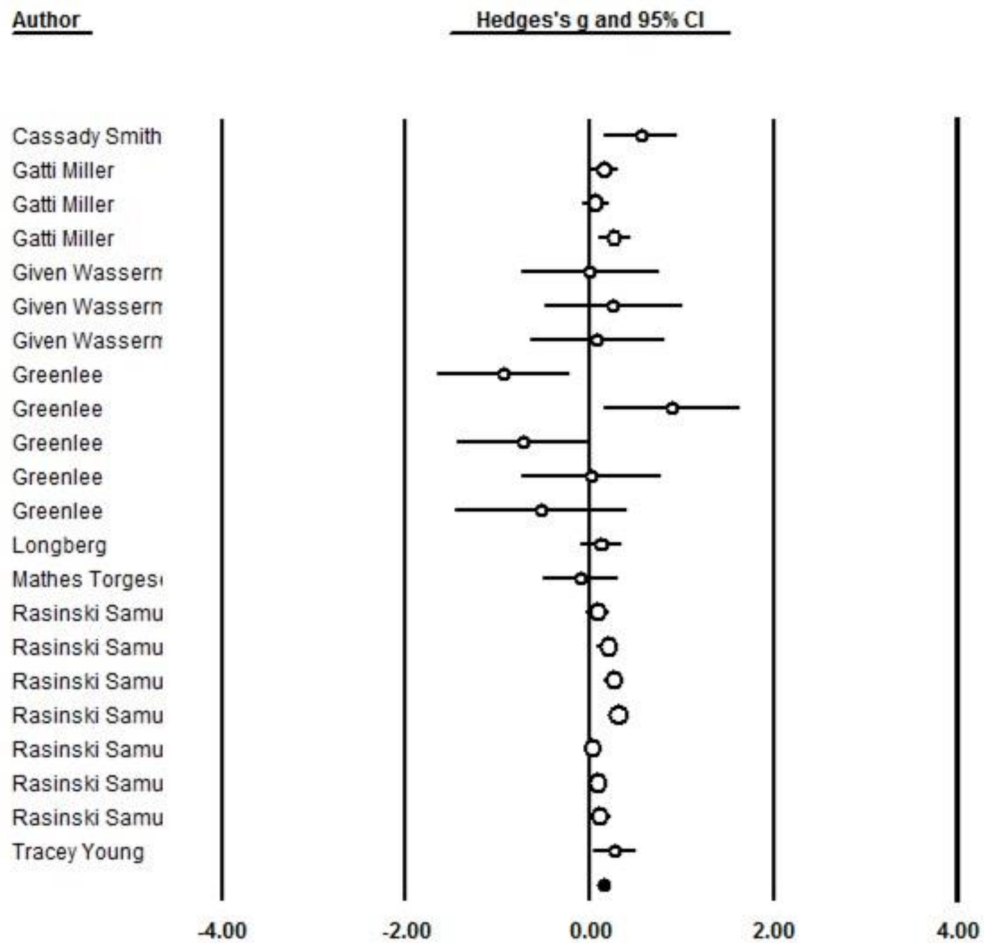

Figure 5. Forest plot of independent effects in Dataset B.

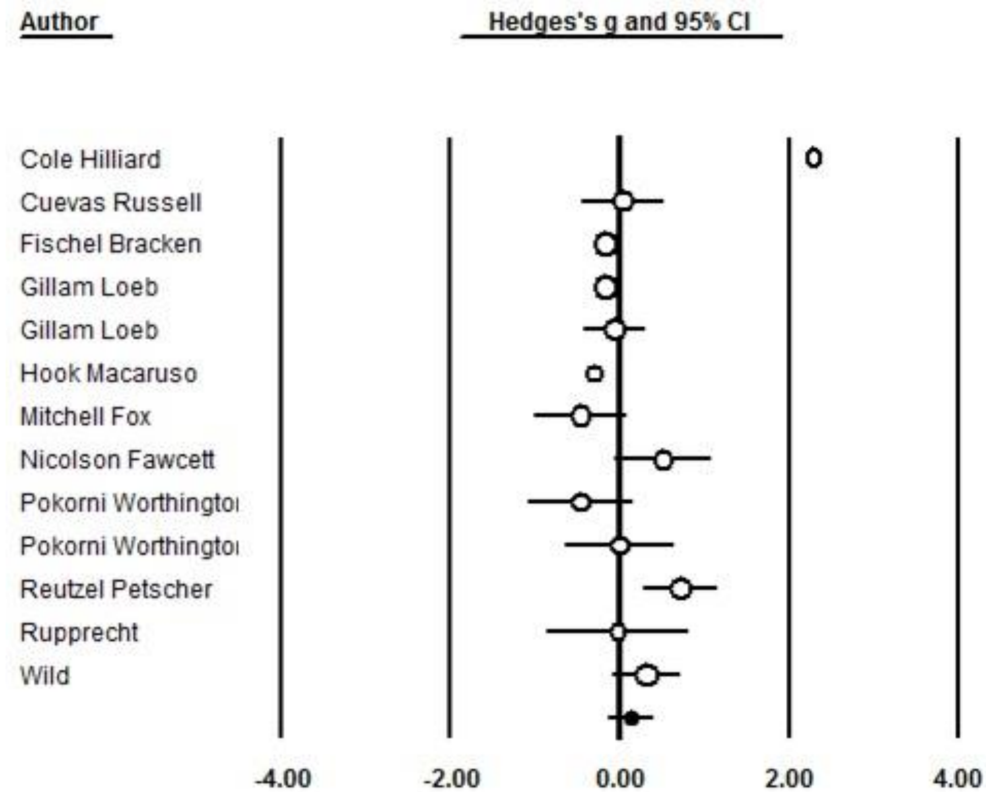

Figure 6. Forest plot of independent effects in Dataset C.

What are the differential effects of CAI on reading outcomes across school levels? To answer this research question, a random-effects analysis was conducted with only independent effects from each report. There is some overlap of these independent effects, such that independent effects from kindergarten are represented across two levels. Table 13 summarizes the analyses at each school level for all three datasets.

In Datasets A and B, the mean effects of preschool through kindergarten were the highest among school levels (g = +0.50 and +0.20, respectively) and were both statistically significant. Studies of students in grades K-6 contributed the most effects across all three datasets and the mean effects were statistically significant in Datasets A and B. The mean effect at the secondary level in Dataset B was larger and statistically significant (g = +0.18) while mean effects at the secondary level in Datasets A and C were near zero. A mean effect from mixed school levels was only reported in Dataset A and was not statistically significant.

Q-values were statistically significant across all school levels except mixed in Dataset A and at the K-6 level in Dataset C. I^2 values ranged from 0%-95% across datasets, with more values in the substantial range in Datasets A and C and more values in the moderate range in Dataset B; thus variability remained between studies within most categories of school level.

In sum, from the available effects, CAI appears to be effective in preschool through elementary school, particularly when compared to a no-treatment control (Datasets A and B), and also appears to be effective at the secondary level when compared to a no-treatment control and assessed using gain scores (Dataset B).

Table 13
Effects of CAI on Reading Outcomes across School Levels

School Level	*k*	*g*	SE	Var	95% CI	*Q*-Value	I^2
Dataset A							
Preschool/Kindergarten	20	+0.50	0.12	0.013	[0.28, 0.73]	151.11*	87%
Elementary (K-6)	50	+0.23	0.09	0.008	[0.06, 0.40]	1024.55*	95%
Secondary (6-12)	9	+0.03	0.08	0.007	[-0.13, 0.18]	33.34*	76%
Mixed	2	-0.22	0.23	0.054	[-0.68, 0.23]	0.04	0%
Dataset B							
Preschool/Kindergarten	2	+0.20	0.09	0.008	[0.03, 0.37]	0.75	0%
Elementary (K-6)	14	+0.12	0.05	0.003	[0.02, 0.22]	28.33	54%
Secondary (6-12)	10	+0.18	0.05	0.002	[0.09, 0.27]	21.15	57%
Mixed	-	-	-	-	-	-	-
Dataset C							
Preschool/Kindergarten	2	+0.04	0.24	0.059	[-0.44, 0.51]	4.24	76%
Elementary (K-6)	10	+0.21	0.20	0.038	[-0.18, 0.59]	46.74*	81%
Secondary (6-12)	2	+0.02	0.22	0.049	[-0.41, 0.46]	0.02	0%
Mixed	-	-	-	-	-	-	-

Note: Only independent effects used in this analysis
*p<0.005

What are the differential effects of CAI on reading outcomes across the five areas of reading? To answer this research question, a random-effects analysis was conducted with all coded, uncombined effects in order to capture the contribution of each dependent variable rather than assessments combined into a single outcome. Analyses including only independent effects were conducted as a test of sensitivity and mean effects were found to be similar if not slightly higher than mean effects using all outcomes; thus, the use of all coded effects presents a slightly more conservative estimate. In addition to the five areas of reading, reports assessing broad reading (i.e., more than one area of reading addressed in one assessment) were analyzed. A variable to capture the effect of aligning the assessments with the areas of reading addressed by the interventions was also analyzed. Table 14 summarizes the analyses for each dataset.

In Datasets A and C, phonemic awareness was assessed most often, and contributed to the highest mean effect in Dataset A (g = +0.33, k = 106). Phonics was also assessed more than other areas of reading in Datasets A and C, though no mean effects from this area of reading were statistically significant in any dataset. Mean effects from assessments including measures of fluency and vocabulary were also not statistically significant across datasets. In Dataset B, comprehension was measured most often, and this mean effect was the highest in this dataset and the only mean effect to reach statistical significance (g = +0.13, k = 102).

Measures of broad reading resulted in statistically significant mean effects across all three datasets (g = +0.28, +0.14, and +0.48, respectively). Measures were aligned with one or more areas of reading addressed by the CAI program more often than not in all three datasets, and these measures led to mean effects reaching statistical significance in Datasets A and B (*g* = +0.27 and +0.13, respectively). Evidence of heterogeneity remained high in Dataset A, with statistically significant *Q*-values and I^2 ranging from 70-99% between studies in all categories. Less heterogeneity remained within categories of Datasets B and C, though statistically significant *Q*-values and I^2 values above 75% indicate remaining heterogeneity within several categories of these datasets.

Table 14

Effects across Reading Outcomes

Outcomes	*k*	*g*	SE	Var	95% CI	*Q*-Value	I^2
Dataset A (*k* = 263)							
Phonemic Awareness	106	+0.33	0.04	0.002	[0.25, 0.41]	347.39*	70%
Phonics	102	+0.24	0.13	0.017	[-0.02, 0.49]	5309.84*	98%
Fluency	20	+0.17	0.09	0.008	[-0.01, 0.35]	62.50*	70%
Vocabulary	24	-0.13	0.31	0.099	[-0.74, 0.49]	1734.46*	99%
Comprehension	38	+0.20	0.12	0.014	[-0.03, 0.43]	587.60*	94%
Broad Reading	46	+0.28	0.05	0.002	[0.18, 0.38]	395.82*	89%
Aligned with Intervention	196	+0.27	0.08	0.006	[0.12, 0.42]	5480.22*	96%
Not Aligned with Intervention	67	+0.05	0.11	0.012	[-0.17, 0.26]	3457.99*	98%
Dataset B (*k* = 217)							
Phonemic Awareness	13	+0.10	0.08	0.007	[-0.06, 0.26]	18.09	34%
Phonics	16	+0.11	0.07	0.004	[-0.02, 0.24]	20.13	25%
Fluency	4	-0.20	0.09	0.008	[-0.19, 0.15]	10.70	72%
Vocabulary	14	+0.05	0.06	0.003	[-0.07, 0.16]	37.45*	65%
Comprehension	102	+0.13	0.02	0.000	[0.09, 0.17]	310.73	68%
Broad Reading	173	+0.14	0.02	0.000	[0.11, 0.18]	797.82*	78%
Aligned with Intervention	205	+0.13	0.02	0.000	[0.10, 0.17]	855.35*	76%
Not Aligned with Intervention	12	+0.05	0.08	0.006	[-0.11, 0.20]	13.42	18%
Dataset C (*k* = 52)							
Phonemic Awareness	21	-0.07	0.07	0.005	[-0.21, 0.06]	31.72	37%
Phonics	18	-0.15	0.10	0.011	[-0.36, 0.05]	49.76*	66%
Fluency	6	+0.24	0.35	0.125	[-0.46, 0.93]	26.80*	81%
Vocabulary	1	-0.21	0.11	0.011	[-0.42, -0.00]	0.00	0%
Comprehension	10	+0.27	0.19	0.038	[-0.11, 0.65]	56.26*	83%
Broad Reading	7	+0.48	0.24	0.059	[0.01, 0.95]	48.97*	88%
Aligned with Intervention	35	+0.05	0.08	0.006	[-0.10, 0.20]	132.34*	74%
Not Aligned with Intervention	17	-0.23	0.07	0.005	[-0.36, -0.09]	19.43	17%

Note: Analyses of outcomes in areas of reading used all coded, uncombined effects.

*$p<0.005$

In sum, effects of CAI compared to a no-treatment control were greater when measures included the areas of phonemic awareness (Dataset A) and comprehension (Dataset B), and when measures were aligned with areas of reading targeted by the CAI program. Broad reading measures, rather than measures of component skills also contributed greater mean effects across datasets.

What methodological characteristics contribute to the effects of CAI in reading? In order to address this research question, categorical moderators were analyzed using a random-effects model including relevant report characteristics. These characteristics were also analyzed using meta-regression to identify variables that help to explain remaining variability among effects. Between-group *Q*-statistics (Q_B) were calculated alongside I^2 as an indicator of heterogeneity between subgroups within each category. Results of all moderator analyses are summarized in Tables E1-E12 located in Appendix E and are separated by datasets according to characteristics of the studies, participants, interventions and CAI programs. In the following paragraphs, I synthesize relevant information gleaned from the categorical moderator analyses and meta-regression for each dataset. When relevant, figures are provided to illustrate relations between effect sizes and specific variables.

In order to be consistent with analyses across datasets containing fewer independent effects, analyses were conducted with all effects in each dataset. Differences between analyses conducted using both independent effects and all effects were not substantial. Of the 217 effects in Dataset B, 160 came from one study (Rasinski et al., 2011).

Study characteristics**.** Six categories of study characteristics were explored, including (a) publication type, (b) publication year, (c) funding source, (d) country of study, (e) location of study, and (f) experimental design (see Tables E1-E3). Effects most often came from journal articles, though results were mixed between Dataset A, where the mean effect from dissertations was higher than journal articles and technical reports (g = +0.39, k = 50), and Dataset B, where the mean effect from journal articles was higher (g = +0.14, k = 193) than effects from other sources. These means and the mean effect from studies published in journals in Dataset A were statistically significant. Mean effects in Dataset C were near zero and not statistically significant. In a meta-regression of publication type on effect sizes, effects in Dataset A from journal articles and dissertations were predicted to be higher than those from technical reports (R^2 = 4.34%, Q_M (2) = 14.97, p = 0.001). I^2 indicated remaining heterogeneity among groups in the category of publication type in Datasets A and B, but was 0% in Dataset C.

Publication year ranged from 2000-2013, with most effects coming from studies made available after 2005. These mean effects were higher and statistically significant across categories, with the exception of those in Dataset C made available before 2010. In Dataset B, a meta-regression revealed that year of publication was a statistically significant predictor of effect size (R^2 = 3.04%, Q_M (1) = 8.06, p = 0.005). As publication year increased, effect sizes also slightly increased (see Figure 7). The Q_B was statistically significant in Dataset B and I^2 ranged from 59%-86% between categories of publication year in all three datasets, indicating remaining heterogeneity between groups.

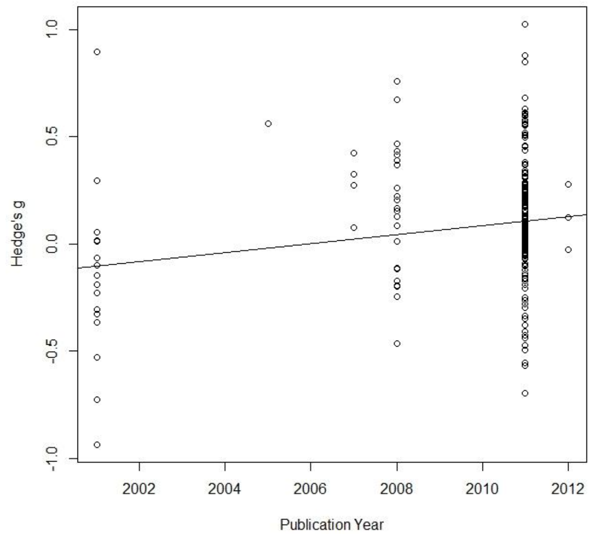

Figure 7. Dataset B: Scatterplot of Hedge's g by publication year

The funding source of studies was most often not specified, and the mean effects in this unspecified category were higher and statistically significant in Datasets A and B. I^2 was 0% across all datasets for the funding source variable. Effects came from studies conducted in the United States more often than other countries; all effects in Dataset B came from studies conducted in the United States. The mean effects of studies in this category were larger and statistically significant in Datasets A. In Dataset C, the five effects from studies conducted in Great Britain contributed to a larger, statistically significant mean effect (*g* = +0.36). There was no heterogeneity between groups in the category of country location in Dataset A, but substantial heterogeneity between groups in Dataset C (I^2 = 85%).

The contribution of study location was varied. Mean effects often came from mixed or unspecified locations in Datasets A and B, and these mean effects (*g* = +0.30

and +0.13, respectively) were statistically significant. I^2 was 10% in this category in Dataset A, indicating increased explanation of heterogeneity between groups. The one effect from a study conducted in a rural area in Dataset B produced a larger, statistically significant effect of +0.56. Five effects from studies conducted in urban areas in Dataset C led to a large, statistically significant mean effect (g = +0.92). Meta-regression of study characteristics in Dataset C revealed that location of study was the only statistically significant predictor and explained a large amount of variation (R^2 = 72.82%, $Q_M(2)$ = 59.39, $p < .001$) with effects from studies conducted in urban areas predicted to be higher than effects from studies conducted in suburban, rural, or mixed/unspecified locations. A model including country and location of study in Dataset C was statistically significant and explained almost 95% of the variation in effects (R^2 = 94.75%, $Q_M(3)$ = 101.64, $p < .001$).

Effects from studies using a randomized control trial (RCT) were more common in Datasets A and C and led to a higher, statistically significant mean effect in Dataset A (g = +0.30). In Dataset A, a meta-regression of experimental design on mean effects explained almost 3% of the variation and a test of the model was statistically significant (R^2 = 2.85%, $Q_M(1)$ = 8.91, p = 0.003). Studies using an RCT design were predicted to have higher effects than those from quasi-experimental designs. Mean effects from both RCTs and quasi-experimental designs in Dataset B were positive, similar and statistically significant. I^2 between groups in this category was 76% in Dataset A, 38% in Dataset B and 0% in Dataset C, indicating reduced heterogeneity between subgroups in the category of experimental design.

***Participant characteristics*.** Four categories of participant characteristics were examined, including (a) total sample size (i.e., large or small); (b) school level; (c) percentage of the sample reported to be at-risk in reading; and (d) percentage of the sample with low socio-economic status, as reported through free-and-reduced lunch (see Tables E4-E6 in Appendix E). Effect size at pretest was also computed and analyzed via meta-regression.

The sample size was categorized into large (n > 250) and small (n < 250) samples. Most of the effects in Dataset A and C came from small samples, and the mean effect of small samples in Dataset A was much higher than the mean from large samples (g = +0.28 vs. -0.19); the mean from small samples was statistically significant. In Dataset B, effects were split relatively evenly among large and small samples, though the mean effect from large samples was higher and statistically significant (g = +0.16). Mean effects from large and small samples in Dataset C were negative and not statistically significant.

Using meta-regression, sample size was examined as a continuous variable rather than a categorical moderator on effects within each dataset. In Dataset A, this variable explained over 4% of the variation, predicting a slight decrease in effect sizes as sample size increased (R^2 = 4.43%, $Q_M(1)$ = 13.99, p < .001). An examination of the scatterplot in Figure 8 reveals wide variation in effects among studies containing around 1000 participants in Dataset A. I^2 between categories of sample size ranged from 86%-91% in Datasets A and B, and Q_B was statistically significant in Dataset B, indicating substantial heterogeneity between groups; in Dataset C, I^2 was much lower (27%).

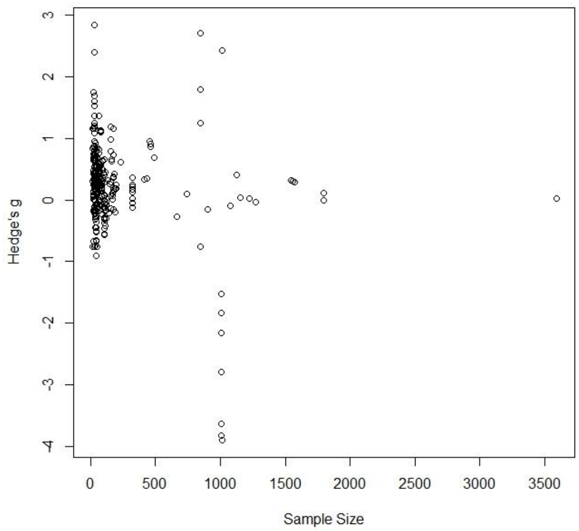

Figure 8. Dataset A: Scatterplot of sample size and Hedge's *g*

In the categorical moderator analysis of school level, levels were recoded to eliminate as much overlap as possible. For example, when categorizing as preschool through kindergarten and kindergarten through 6th grade, kindergarten was represented within two categories. By recoding the variables as (a) preschool through elementary (i.e., grades PreK-6); (b) middle school (i.e., grades 6-8); (c) high school (i.e., grades 9-12); and (d) mixed or unspecified, the overlap across levels was eliminated. Though sixth grade appears to be represented in two categories, studies were separated by those including sixth grade as elementary participants or as middle school participants (i.e., grades 6-8); thus, samples did not overlap.

In Datasets A and C, most effects came from studies conducted at the preschool through elementary level. This mean effect in Dataset A was larger than other school levels with the exception of one effect at the high school level, and was the only category mean to reach statistical significance in this dataset (g = +0.24). In Dataset B, more effects came from studies conducted at the middle school level and this mean effect was higher than the mean effects at other levels (*g* = +0.19). Mean effects across all school levels in Dataset B were positive and statistically significant. The bulk of effects in Dataset B came from one study (Rasinski et al., 2011), and effects from this study are represented across all school levels. School level did not appear to impact the results of Dataset C, as all mean effects were near zero and not statistically significant. Categorization across school level resulted in decreased heterogeneity in Datasets A and C (I^2 = 0%) but heterogeneity remained substantial in Dataset B (I^2 = 85%).

Percentages of the sample reported to be at-risk in reading were categorized as 0-33%, 34-66%, 67-100%, and not specified. In Datasets A and B, most effects came from studies not specifying the at-risk status of the participants, and both means were statistically significant (*g* = +0.26 and +0.14, respectively). Many effects across all three datasets came from studies reporting over two-thirds of their samples as at-risk in reading, though these mean effects were not statistically significant. I^2 was low across all three datasets, indicating less heterogeneity between mean effects when grouped by at-risk status. Socioeconomic status was typically not reported across datasets; however the mean effects from studies reporting over two-thirds of their samples with low SES in Datasets A and B were both positive and statistically significant (*g* = +0.54 and +0.13, respectively). This moderator reduced I^2 to low to moderate levels across all three datasets.

For studies providing pretest data, effect sizes were computed using the same methods to compute posttest effect sizes. A positive pretest effect size would indicate that pretest scores of the treatment group were higher than pretest scores of the control group. Pretest scores were provided for 201 effects in Dataset A and explained over 13% of the variation in effects (R^2 = 13.88%, $Q_M(1) = 22.85$, $p < .001$), indicating higher posttest effects with increased magnitude in pretest effects. In Dataset B, pretest effect sizes were computed for 45 of the 217 effects. This moderator was not statistically significant and did not account for any portion of the variation in effects. In Dataset C, pretest effect sizes were provided for 47 of the 52 effects and accounted for over 27% of the variation in effects (R^2 = 27.28%, $Q_M(1) = 8.72$, $p = 0.003$). Similar to Dataset A, this model indicated that as pretest effect sizes increased, posttest effect sizes also increased.

Intervention characteristics. Six categories of intervention characteristics were explored, including (a) location of the intervention (e.g., core literacy classroom or separate room); (b) the literacy component (e.g., supplemental to core literacy or integrated); (c) the session length in minutes; (d) the number of sessions per week; (e) the total duration of the intervention in hours; and (f) the type of teacher leading the intervention (e.g., licensed teacher or researcher). Tables E7-E9 in Appendix E provide a summary of the categorical moderator analyses.

Effects came from studies providing interventions in separate rooms most often across all three datasets. In Datasets A and B, these groups contributed to a statistically significant mean effect (g = +0.31 and +0.09, respectively). In Dataset B, the mean of the four effects from studies conducted within core literacy classrooms was higher and statistically significant (g = +0.32). A meta-regression of intervention location on effect sizes explained almost 5% of the variation in effects in Dataset B (R^2 = 4.81%, Q_M (3) =

11.07, p = 0.011) indicating higher effects for interventions located in core literacy classrooms. Similarly, the mean effect of CAI integrated into core literacy instruction was higher and statistically significant in Dataset B (g = +0.13), as was the mean effect from studies with a mixed or unspecified literacy component. The mean effect of CAI supplemental to core instruction was higher and statistically significant in Dataset A (g = +0.29). Effects in the categories of intervention location and literacy component were near or below zero in Dataset C and did not reach statistical significance. There was little remaining heterogeneity between mean effects in Dataset A (I^2 = 0%) and moderate heterogeneity in Datasets B and C, ranging from 57%-69% between groups when studies were categorized by intervention location and literacy component.

Session length was grouped by sessions lasting up to 30 min, those lasting 30 min to 1 hr, and those lasting longer than one hr. Session length varied across datasets, as did the mean effects between groups. In Dataset A, more effects came from interventions lasting up to 30 min per session, and this mean (g = +0.25) was statistically significant. I^2 was 0%, indicating no heterogeneity between these groups. In Dataset B, session length was most often between 30-59 min, and the mean effects from sessions lasting 1-29 min or 30-59 min were similar, positive, and statistically significant (g = +0.12 and +0.13, respectively). In Dataset C, more effects came from studies meeting for over an hour per session, but the only mean effect to reach statistical significance came from three effects with sessions lasting 30-59 min (g = +0.66). I^2 was 72% in Dataset B and 73% in Dataset C, indicating remaining heterogeneity between groups.

The number of sessions per week was analyzed by grouping effects by those from interventions meeting once or twice per week, three to four times per week, five times per week, or those not specifying the number of sessions. In Datasets A and C, most interventions met five times per week; both mean effects were near or below zero and not statistically significant. Effects from interventions meeting once or twice a week led to higher, statistically significant means in Datasets A and C (*g* = +0.29 and +0.55, respectively), though the mean for studies not specifying how many sessions occurred per week was higher and statistically significant in Dataset A (*g* = +0.47). Effects from interventions meeting three to four times per week were more common in Dataset B (*g* = +0.14); this and the mean effect from interventions meeting five times per week (*g* = +0.22) were statistically significant. In Dataset C, the mean effect from studies meeting three to four times per week was statistically significant (*g* = +0.34). Meta-regression of this variable in Dataset C revealed that as the number of sessions per week increased, effects decreased (R^2 = 22.77%, Q_M (1) = 9.95, p < .001). I^2 was moderate to substantial between groups across datasets in the category of sessions per week, ranging from 63% to 93%.

The total duration of intervention was grouped by those meeting up to 10 hours, 11-20 hours, 21-50 hours, 51-100 hours, or those with unspecified durations. Duration varied across datasets, but most effects came from interventions lasting up to 50 hours total. Larger mean effects came from studies with a total duration up to 10 hours in Datasets A and C (*g* = +0.43 and +0.23, respectively), and both means were statistically significant. Positive, statistically significant mean effects were found for studies with durations of 11-20 hours in Datasets A and B (g = +0.30 and +0.19, respectively), and the largest mean effect in Dataset B came from the group of effects with durations of 21-50 hours (g = +0.32). This mean effect was also statistically significant. In Dataset C, effects from studies meeting 21-50 hours total contributed to a statistically significant negative mean effect (g = -0.22).

In the meta-regression analyses, total duration consistently explained variation in effects across all three datasets. In Dataset A, as duration increased, effects decreased (see Figure 9; R^2 = 2.55%, Q_M (1) = 9.52, p = 0.002). The opposite was found in Dataset B, where increased duration led to slightly increased effects (see Figure 10; R^2 = 3.14%, Q_M (1) = 3.65, p = 0.055). Results from Dataset C were similar to Dataset A, revealing decreased effects with increased duration (see Figure 11; R^2 = 9.80%, Q_M (1) = 6.81, p =0.009). I^2 was moderate to substantial in this category across datasets, ranging from 62%-85%, indicating moderate to substantial heterogeneity among groups in this category across datasets.

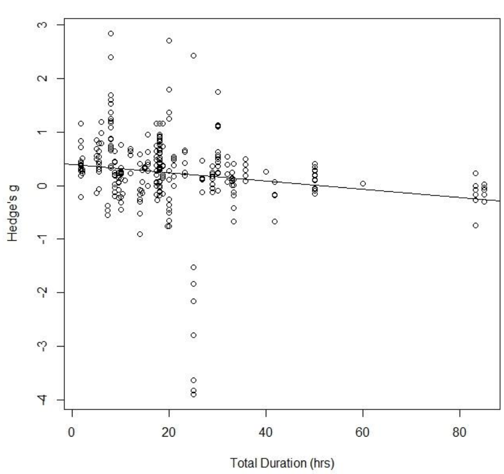

Figure 9. Dataset A: Scatterplot of Hedge's g by total duration (hrs)

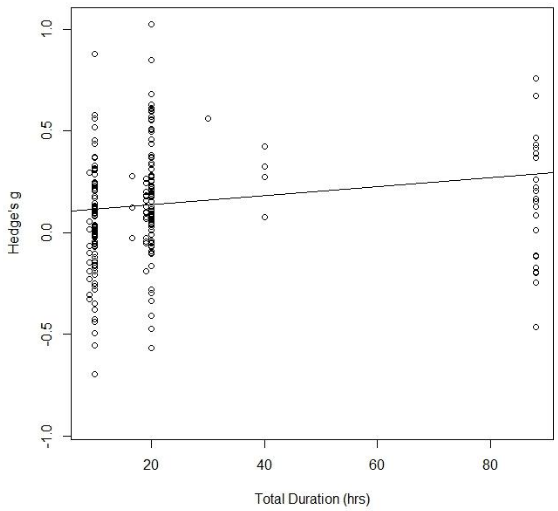

Figure 10. Dataset B: Scatterplot of Hedge's *g* by Total Duration (hrs)

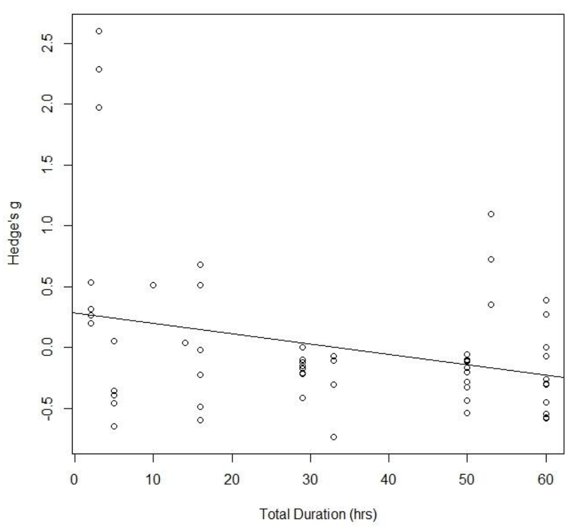

Figure 11. Dataset C: Scatterplot of Hedge's *g* by Total Duration

The type of teacher implementing the intervention (i.e., the adult monitoring students' use of CAI) was categorized as licensed teachers, researchers, other school staff (e.g., teaching assistants or volunteers), or unspecified personnel. Licensed teachers were most often represented in effects from Datasets A and C. In all three datasets, effects from CAI led by licensed teachers were statistically significant; however, the mean effects from Datasets A and B were positive (g = +0.17 and +0.13, respectively) while the mean effect from Dataset C was negative (g = -0.19). In Datasets A and C, mean effects from CAI led by researchers was positive and statistically significant (g = +0.55 and +0.47, respectively). In Dataset B, most effects came from studies not specifying personnel involved in implementation, and this mean effect as well as the mean from two studies in Dataset C, was positive and statistically significant (g = +0.14 and +0.71, respectively).

In the meta-regression analyses of intervention characteristics, models including teacher type helped explain variation across all three datasets. In Dataset A, the type of teacher implementing the intervention explained over 3% of the variation in effects, with higher effects predicted from researchers and licensed teachers (R^2 = 3.44%, Q_M (3) = 15.38, p = 0.002). In Dataset B, the type of teacher was a statistically significant predictor of effect (R^2 = 6.09%, Q_M (3) = 15.48, p = 002), with interventions led by researchers predicted to have reduced effects compared to other personnel. In Dataset C, the type of teacher explained over 80% of the variation in effects; licensed teachers were found to have lower effects than researchers or unspecified personnel (R^2 = 81.71%, Q_M (2) = 53.48, $p < .001$).

In Dataset B, a model including the location of the intervention, the literacy component, and total duration of the intervention explained the greatest amount of variability in effects (R^2 = 15.5%, Q_M (5) = 28.52, $p < .0001$) and predicted higher effects for studies with a mixed or unspecified literacy component, conducted in mixed or separate locations, with increased total duration. In Dataset C, a model including teacher type, total duration, sessions per week, the literacy component, and the intervention location accounted for almost 90% of the variation in effects (R^2 = 88.37%, $Q_M(7)$ = 87.91, $p < .001$); however a second model eliminating the teacher type accounted for a similar amount of variation in effects (R^2 = 89.88%, $Q_M(6)$ = 85.41, $p < .001$). Within this model, the coefficient representing sessions per week was positive, predicting higher effects for studies meeting more often each week when total duration was lower and the literacy components and intervention locations were mixed or not specified.

CAI characteristics. Three main categories of CAI characteristics were examined, including whether or not the CAI program was commercially available, linear or adaptive in structure, and the types and number of instructional design components included (i.e., guided activity, corrective feedback, elaborate feedback, pacing, pretraining, and record keeping; see Tables E10-E12 in Appendix E). Across all three datasets, effects most often came from commercially-available programs, and the mean effects in Datasets A and B were positive and statistically significant (g = +0.21 and +0.13, respectively). In Dataset C, six effects from non-commercially available programs contributed to a positive, statistically significant effect (g = +0.31). I^2 was 0% in Dataset A, indicating no heterogeneity between groups of this variable. In Dataset B, only commercially available programs were included. In Dataset C, there was evidence of substantial heterogeneity between groups in the category of commercial availability (I^2 = 84%).

The structure of CAI did not impact mean effects in Dataset C. In Dataset A, most effects came from studies using CAI that was linear in nature (i.e., progressed sequentially regardless of student input), but only effects from adaptive programs (i.e., program changes sequence based on student input) reached statistical significance (g = +0.34). This result was also found in Dataset B, where the 160 effects from the Rasinski et al. (2011) study contributed a mean of +0.15 for a program that adapted to student input. Other mean effects in Datasets A and B were not statistically significant. Across all three datasets, I^2 ranged from 0%-40%, indicating less heterogeneity when the structure of CAI was taken into account.

The most commonly reported instructional design components across all three datasets were corrective feedback, pacing, and record keeping. Inclusion of each of these three components contributed to higher, statistically significant mean effects in Datasets A and B. In Dataset C, inclusion of record keeping contributed to a negative, statistically significant mean effect (g = -0.14). Though less commonly reported, guided activity contributed to higher, statistically significant mean effects in Datasets A and C (g = +0.34 and +0.65, respectively). Studies that did not report pretraining (i.e., activating background knowledge) contributed to a higher, statistically significant mean effect in Dataset A (g = +0.24) while the two effects from studies including pretraining in Dataset C resulted in a large, statistically significant mean effect (g = +2.26). Studies not reporting elaborate feedback contributed to a higher, statistically significant mean effect in Dataset A (g = +0.23).

The number of instructional design components included was separated into categories of zero to one component, two to three components, and four to six components. Across all three datasets, most studies reported two to three components. In Dataset B, this group contributed to a higher, statistically significant mean effect (g = +0.15). In Dataset A, however, only effects from studies reporting zero to one component or four to six components contributed to statistically significant mean effects (g = +0.33 and +0.29, respectively).

In Dataset A, I^2 ranged from 0%-67% among instructional design components, indicating less heterogeneity between groups in each category. A meta-regression model including guided activity and pretraining resulted in the most explained variation in effects in Dataset A, with programs including guided activity but not reporting pretraining predicted to have increased effects (R^2 = 2.85%, Q_M (2) = 11.47, p = 0.003). Meta-regression of the instructional design components in Dataset B revealed that the number of components was a statistically significant predictor of effect (R^2 = 4.86%, Q_M(1) = 8.09, p = 0.005). A model including only record keeping as a moderator explained more variability and was statistically significant (R^2 = 6.37%, Q_M (1) = 14.12, $p < .001$) indicating that studies including effects with record keeping resulted in increased effect sizes. In Dataset C, guided activity and elaborate feedback were the only statistically significant coefficients. A model including these two moderators accounted for over 70% of the variability in effects (R^2 = 72.19%, Q_M (2) = 68.52, $p < .001$). Mean effects from studies reporting guided activity were predicted to be higher, while mean effects from studies reporting elaborate feedback were predicted to be lower.

Meta-regression of methodological characteristics. Models including categorical moderators and continuous variables that helped to explain variation in effects across categories of methodological characteristics were tested. In Dataset A, a model including pretest effect size, experimental design type, and teacher type accounted for almost 30% of the variation in effects (R^2 = 29.17%, $Q_M(5) = 53.86$, $p < .001$) and indicated that effects were predicted to be higher when researchers or licensed teachers led the interventions, when the pretest effect size was higher, and when the studies utilized a quasi-experimental design.

In Dataset B, the previously mentioned model including intervention location, literacy component, and total length explained the greatest amount of variation in effects (R^2 = 15.50%, $Q_M(5) = 28.52$, $p < .001$) and indicated that studies with a mixed or unspecified literacy component, conducted in a mixed or separate location, and with increased total duration were predicted to result in higher effects. In Dataset C, a model including pretest effect size, the location of the study, and the location of the intervention accounted for 100% of the variability (R^2 = 100%, $Q_M(6) = 103.01$, $p < .001$) and indicated that studies conducted in urban areas, in mixed intervention locations, and with higher pretest effect sizes were predicted to have higher effects.

In sum, the moderator analyses and meta-regression revealed that several variables influenced mean effects across datasets. Pretest effect sizes influenced mean effects in two of three datasets, and the year the study was made available influenced mean effects in all three datasets. Intervention characteristics, including the location of the intervention, the literacy component, number of sessions per week, and total duration also played a role in moderating mean effects across datasets.

What is the overall quality of the research according to the quality indicators proposed by Gersten et al. (2005)? To answer this research question, each study was coded according to the quality indicators suggested by Gersten et al. (2005), including participant description, intervention and control description, outcome measures, data analysis, research design, and overall quality score. Table 4 (p. 59) provides a description of each quality indicator. A total of 61 studies were coded in this meta-analysis, with seven studies represented in both Datasets A and C. Table F1 in Appendix F provides a summary of the quality indicators met by each study. The independent effects were analyzed via categorical moderator analyses to assess the impact each variable had on the mean effect of each dataset. Tables G1-G3 in Appendix G display the effects of the categorical analyses of research quality.

Across all studies, studies met between 17% and 92% of quality indicators (M = 54%, *Mdn* = 58%), with the indicators of intervention description (85%), measures timed appropriately (92%), and data analysis (89%) most prevalent. Indicators with a lower percentage met included comparable interventionists (33%), attrition less than 30% (23%), fidelity (21%), and multiple measures (33%). The quality indicators of ability/disability description, comparable groups, control description, reported effect sizes, and RCT were met in about half of all studies.

Dataset A. Dataset A consisted of 49 studies contributing 66 independent effects. Overall, 52% of the quality indicators were met in this dataset. Relative strengths among indicators included the description of the intervention (90% of studies), the timing of the measures (94%) and an appropriate data analysis (90%). Over half of the independent effects came from studies not adequately describing participants, but with sufficient evidence of group comparability. Attrition and evidence of comparable interventionists were not provided in most of the effects in Dataset A. For studies that did provide evidence of comparable interventionists and multiple outcome measures, the mean effect was relatively higher than other categories (k = +0.41 for both categories). Over half of all studies failed to adequately describe the control condition. Most studies did not employ multiple measures and use of an RCT was met in over half of the studies in Dataset A.

In Dataset A, a meta-regression model including the moderators of control description, multiple measures, and RCT explained over 8% of the variability in effects (R^2 = 8.87%, $Q_M(3) = 31.84$, $p < .001$). The control description coefficient was negative, suggesting that studies reporting an adequate description of the control condition were predicted to have lower effects when studies used an RCT and assessed outcomes with multiple measures.

Dataset B. Dataset B consisted of eight studies contributing 22 independent effects. Overall, 59% of quality indicators were met by studies in Dataset B. Relative strengths among indicators included the description of abilities/disabilities (75%), the intervention description (75%), timing of the measures (88%), and an appropriate data analysis (88%). The lowest percentages of indicators met included comparable groups (38%) and fidelity of implementation (13%). Mean effects were similar among all categories except data analysis, where five effects from studies not reporting an appropriate analysis resulted in a negative mean effect.

In Dataset B, a model including participant ability descriptions, comparable groups, comparable interventionists, and attrition explained over 5% of the variation in effects (R^2 = 5.42%, $Q_M(4) = 14.17$, $p = 0.007$). The coefficient representing ability/disability description was negative, suggesting that studies adequately reporting the participants' abilities were predicted to have lower effects. The other three coefficients were positive.

Dataset C. Dataset C consisted of 11 studies contributing 13 independent effects. Seven of these studies were also represented in Dataset A. Overall, 59% of quality indicators were met. Relative strengths among indicators included the description of the control condition, the timing of the measures, and an appropriate data analysis (91% for each indicator). The indicators of fidelity (18%), attrition (27%), multiple measures (36%), and reported effect sizes (36%) were relatively less prevalent across this set of studies. The only mean effect to reach statistical significance was from four studies with multiple measures ($g = +0.58$). This was also the largest effect among the categories.

In Dataset C, a model including comparable interventionists, attrition, multiple measures, and measures timed appropriately explained over 96% of the variation in effects (R^2 = 96.47%, $Q_M(4) = 128.06$, $p < .001$). The coefficient for comparable interventionists was negative, suggesting that effects from studies providing evidence of comparability among interventionists were predicted to be lower. The other three coefficients were positive.

Effects by quality score. In order to capture effects across levels of research quality, effects were grouped by percentage of their quality score. Mean effects in each subgroup of Dataset C were not statistically significant. In Datasets A and B, most studies met between 26%-50% of the quality indicators. The mean effects from these studies were positive and statistically significant (*g* = +0.26 and +0.14, respectively). In Dataset B, the mean effect from reports meeting 51%-75% of the indicators was also positive and statistically significant (*g* = +0.17). Though many reports met 51%-75% of the indicators in Dataset A, the mean effect was not statistically significant. Relatively few reports met over 75% of the quality indicators across datasets, and no mean effects in this subgroup were statistically significant.

In sum, studies with quality scores near 50% tended to contribute to higher mean effects in Datasets A and B. Several common variables appeared to impact mean effects across at least two datasets, including the use of multiple measures, the reporting of attrition, and the comparability of the interventionists. The influence of other quality indicators varied within each dataset.

Publication Bias

Publication bias was examined by generating funnel plots to examine the symmetry of effect sizes centered around the mean effect. Using independent effects, each dataset was analyzed separately.

Dataset A. The funnel plot for Dataset A can be viewed in Figure 12. The plot shows effects distributed fairly evenly around the mean with a few outliers toward the top of the figure. Three studies toward the bottom of the funnel are clustered to the right of the mean, indicating some bias in small studies reporting relatively larger effect sizes. Removal of the outliers below and above the mean (Gale, 2006; Hansen, Llosa, & Slayton, 2004; Huffstetter, King, Onwuegbuzi, Schneider, & Powell-Smith, 2010) did not substantially change the symmetry of the plotted points. The classic fail-safe N indicated an additional 1,070 'null' studies would need to be located for the effect to be nullified. By categorizing independent effects into journal articles vs. dissertations or technical reports, the meta-analysis revealed effect sizes of +0.30 (*k* = 38) and +0.17 (*k* = 28) respectively.

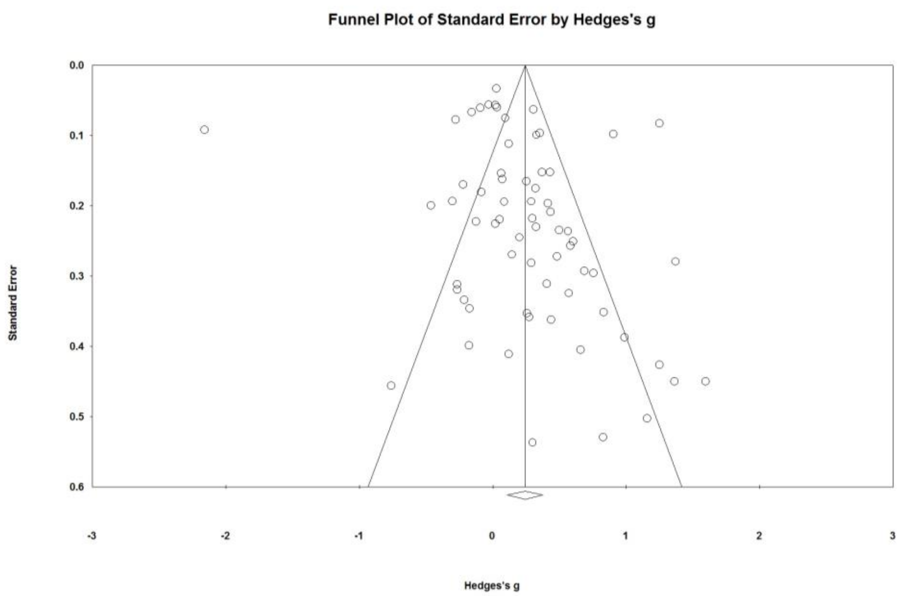

Figure 12. Funnel plot of all independent effects in Dataset A

Dataset B. The funnel plot for Dataset B can be viewed in Figure 13. The plot shows some asymmetry around the mean, with two potential outliers to the left of the mean. The plot was reexamined after removal of one independent effect from Greenlee (2001), a dissertation contributing an effect size of -0.94. The resulting plot was relatively unchanged in terms of symmetry. The classic fail-safe N indicated 178 additional 'null' studies would need to be located to nullify the effect. By categorizing independent effects into journal articles vs. dissertations or technical reports, the meta-analysis revealed effect sizes of +0.14 (k = 9) and +0.06 (k = 13) respectively.

Dataset C. The funnel plot for Dataset C can be viewed in Figure 14. The plot shows effect sizes distributed symmetrically around the mean at the top of the funnel, with two effects to the left of the mean and one large positive effect toward the bottom of the funnel. Removal of the extreme positive effect represented in Cole and Hilliard (2006; g = +2.29) improved the symmetry of the funnel. Since the combined effect was not statistically significant, the use of the fail-safe N to address spurious significance was not relevant. By categorizing independent effects into journal articles vs. dissertations or technical reports, the meta-analysis revealed effect sizes of -0.05 (k = 12) and -0.02 (k = 1) respectively.

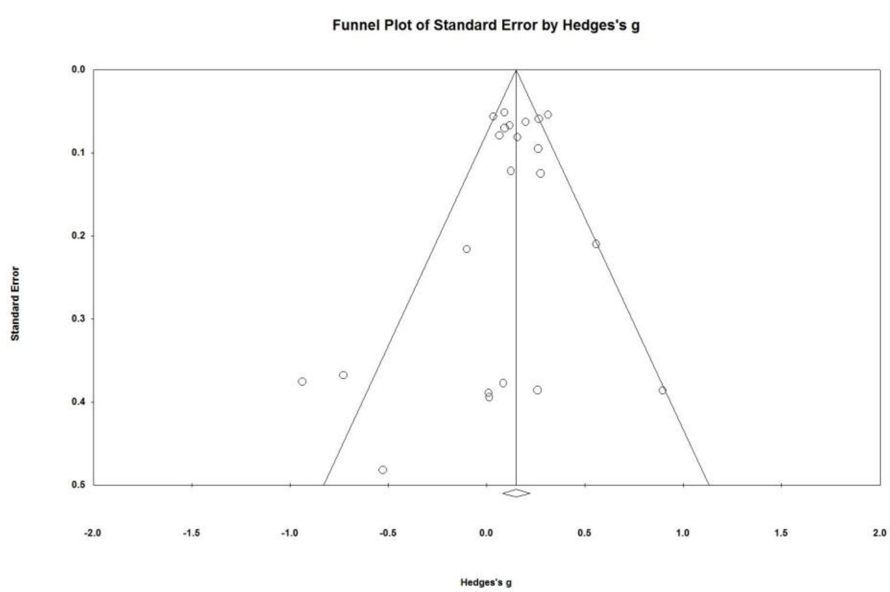

Figure 13. Funnel plot of all independent effects in Dataset B

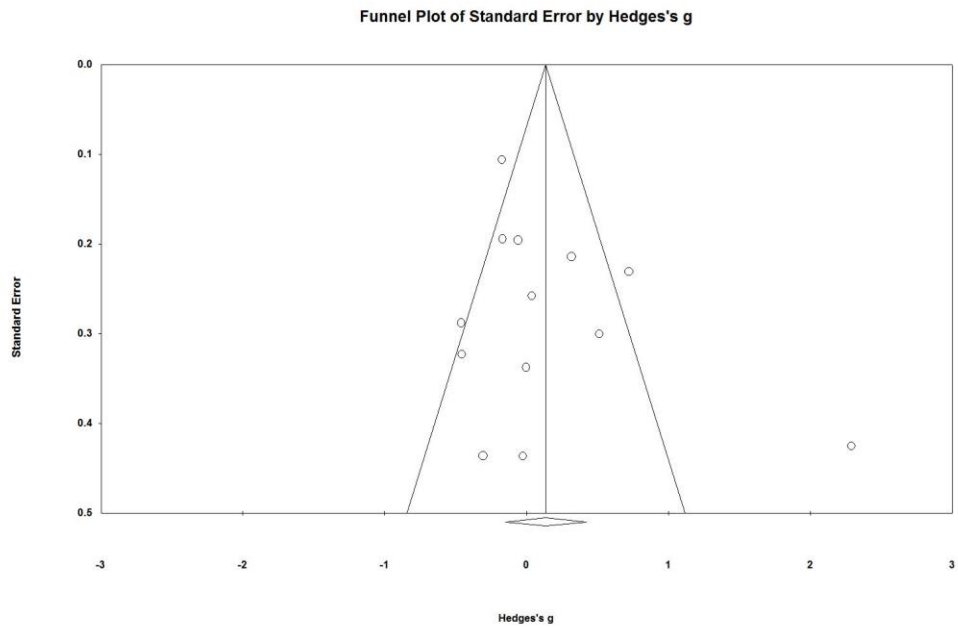

Figure 14. Funnel plot of all independent effects in Dataset C

Chapter 5

DISCUSSION

The primary purpose of this dissertation was to examine the effects of computer-assisted instruction (CAI) on reading outcomes for students in preschool through high school. Using meta-analytic techniques (Cooper et al., 2009), effects of studies investigating CAI in reading were synthesized and examined for evidence of variability across (a) school levels (preschool, elementary, middle, and high school); (b) areas of reading (phonemic awareness, phonics, fluency, vocabulary, comprehension); and (c) study characteristics (e.g., methodological characteristics and instructional design components). In addition, an examination of research quality was conducted to help determine future research needs and analyze the effects of research quality on reading outcomes.

In this chapter, I provide a summary and discussion of the results presented in Chapter 4. I compare findings from this meta-analysis with findings from previous research, and discuss practical implications. Limitations of this meta-analysis and suggestions for future research are also discussed.

Summary and Interpretation of Results

A search of the literature base on the effects of CAI in reading yielded 61 studies that met criteria designed to limit the analysis to studies investigating CAI as defined in this study, rather than those investigating assistive technology (AT) or broad educational technology (ET). This analysis focused on outcomes in reading for students in preschool through high school and, in order to capture the most current research available, only studies made available from 2000 were included. Across all three datasets, 532 effects sizes were extracted, 101 of which were independent effects. In the following paragraphs I summarize and interpret the results according to each research question.

What is the effect of CAI on reading outcomes for students in preschool through high school and what is the evidence of variability among those effects? The three datasets examined in this dissertation included studies comparing (a) CAI with no treatment or non-reading CAI reporting posttest scores (Dataset A); CAI with no treatment reporting gain scores (Dataset B); and (c) CAI with traditional, teacher-led intervention reporting posttest scores (Dataset C). Cohen (1988) offered an interpretation of the magnitude of effect sizes in social science research, describing +0.20 as small, +0.50 as moderate, and +0.80 as large. Summary effects indicate that CAI had a small, positive effect on reading outcomes for students in preschool through high school when compared with peers receiving non-reading CAI or no treatment.

The mean effect was slightly higher for studies reporting posttest scores (k = 66, g = +0.24) than gain scores (k = 22, g = +0.15). With fewer independent effects measured with gain scores, small or negative effect size estimates may have had more influence on the mean effect in Dataset B than in Dataset A. For example, 18 of the 22 independent effect sizes came from just four reports. One report (Greenlee, 2001) contributed five independent effects, two of which were large, statistically significant, negative effect sizes (g = -0.94 and -0.73). Two other large studies (i.e., >250 participants) contributed 10 independent effect sizes (Gatti & Miller, 2011; Rasinski et al., 2011), and most of those effects were very small. The relative weights of these studies were much larger than those of studies with fewer participants, suggesting greater influence on the summary effect. In contrast, study weights were relatively even in Dataset A, suggesting less influence of large studies on the summary effect. Thus, individual studies may have had a stronger influence on independent effects in Dataset B than in Dataset A.

Though the mean of the independent effects in Dataset C was positive and similar to that of Dataset B (k = 13, g = +0.14), the summary effect was not statistically significant. Furthermore, the removal of Cole and Hilliard's (2000) outlier effect reduced the mean of the independent effects to +0.01. The results from Dataset C indicate that the reading outcomes for students receiving CAI were not different from the outcomes for students receiving teacher-led intervention; in other words, increases or decreases in reading outcomes were similar for both groups, regardless of the instructional delivery system. Given that reading intervention research has reported positive outcomes for students receiving targeted, teacher-directed reading interventions (e.g., Swanson, 2012; Scammacca et al., 2007; Swanson, 1999), reading scores would be expected to increase for both groups, resulting in effect sizes near zero.

Across all three datasets, Q-values were statistically significant and I^2 values were moderate to large, indicating variation due to between-studies heterogeneity (Higgins & Thompson, 2002; Huedo-Medina et al., 2006). In an effort to account for this variability among effects, I categorized effects by school level, reading outcomes, methodological characteristics, and research quality. Below, I explore differential effects by these variables.

What are the differential effects of CAI on reading outcomes across school levels? Results suggest that students in preschool through elementary school benefitted from CAI when compared to students receiving no treatment or non-reading CAI. Results indicated no difference between CAI and teacher-led intervention across school levels. (Dataset C). When only gain scores were reported (Dataset B), students in secondary schools obtained positive effects similar to those found in preschool through elementary school. Dataset B had a higher proportion of effects from the secondary level than the other datasets, and the two large studies previously mentioned (Gatti & Miller, 2011; Rasinski et al., 2011) contributed larger effects with relatively greater weights, warranting further investigation of those studies and the represented CAI programs.

Heterogeneity was reduced in Datasets A and C when school level was examined as a moderator (Higgins et al., 2003), suggesting that mean effects between levels were more consistent than ungrouped effects. In Dataset B, effects from students in middle school (i.e., grades six through eight) were relatively higher than those at other school levels, though many of the non-independent effects making up the mean for students in middle school came from one study (Rasinski et al., 2011).

Overall, these results indicate that CAI is effective across all school levels when compared to no treatment or non-reading CAI. It is encouraging to find positive effects for students at the secondary level, as the bulk of research on CAI in reading has been conducted at the preschool through elementary level. These results align with a meta-analysis examining interventions for adolescent struggling readers (Scammacca et al., 2007), concluding that secondary students can and do benefit from targeted interventions in reading.

What are the differential effects of CAI on reading outcomes across the five areas of reading? For this research question, all independent and non-independent effects were included to capture the multiple assessments used in several studies. There was significant overlap in areas of reading, as most effects came from assessments including more than one area of reading. Results suggest that outcomes including phonemic awareness (Dataset A) and comprehension (Dataset B) resulted in small, positive, statistically significant mean effects. Mean effects from outcomes including phonics and fluency were not statistically significant across datasets, which was surprising given the large number of studies conducted at the preschool through elementary level. However, the lower limit of the confidence intervals for outcomes including phonics was -0.02 in both Datasets A and B and -0.01 for fluency in Dataset A, suggesting positive mean effects approaching statistical significance. Measures including vocabulary consistently contributed to small, nonsignificant mean effects. Relatively few CAI programs addressed vocabulary explicitly, suggesting a lack of evidence in the field regarding the use of CAI to target outcomes in this area of reading.

Given the higher proportion of participants at the preschool through elementary level in Dataset A and at the secondary level in Dataset B, preliminary conclusions align with the stages of reading development described in Chapter 2. Chall (1983) suggested that the first two stages of reading development, which typically encompass students in preschool through third grade, focus on phonemic awareness and decoding, so findings of positive effects in these areas of reading for a higher proportion of elementary participants suggests that interventions targeting students' reading development can effectively enhance early reading outcomes. Similarly, Chall (1983) suggested that stages three and four, which typically encompass students in fourth grade through high school, focus on comprehension. Results from this meta-analysis indicate that comprehension can be effectively targeted by CAI when implemented with a larger proportion of secondary students, suggesting that interventions targeting more advanced reading development with older students can also enhance outcomes in reading.

Measures of broad reading (i.e., including two or more areas of reading) resulted in small to moderate, statistically significant effects across all three datasets (g = +0.28, +0.14, and +0.48, respectively). In Dataset C, this moderate mean effect was a relative anomaly. Only seven effects made up this mean, and a statistically significant Q-value and I^2 of 88% suggest unexplained variation between these seven effects that could be further explored. Mean effects from measures that were aligned with one or more areas of reading addressed by the CAI program were small, positive, and statistically significant in Datasets A and B. These findings suggest that outcomes from measures including two or more areas of reading that are also aligned with the reading focus of the CAI being used may be higher than from measures of singular areas of reading not aligned with the area(s) of reading being targeted by the CAI.

When studies reported more than one area of reading per outcome, effects from one area (e.g., comprehension) may have also been included in another area (e.g., vocabulary). Therefore, it is difficult to draw definitive conclusions regarding greater benefits of CAI in one area of reading over another. Given these results, however, educators can be confident that CAI can enhance the early reading outcomes of elementary students and the comprehension outcomes of secondary students.

What methodological characteristics contribute to the effects of CAI in reading? Study, participant, intervention, and CAI characteristics were entered into meta-regression and categorical moderator analyses to investigate differences in effects, potentially explain remaining variability among effects, and provide educators with information regarding factors that may affect implementation outcomes.

Study characteristics. Most of the 61 studies reported in this synthesis came from published journals and were conducted in the United States. Mean effects from studies published in journals were more prevalent and generally higher, though the results of this group varied across datasets. Dataset A had a relatively higher concentration of effects from dissertations, and this mean effect (+0.39) was moderate and statistically significant. This finding was surprising, as peer-reviewed studies tend to have higher effects than unpublished studies, often resulting in publication bias (Lipsey & Wilson, 2001); however, these results may encourage future meta-analysts to include dissertations among the base of evidence for any given practice.

The year the study was made available seemed to play a role in outcomes, as studies conducted after 2005 in Datasets A and B and after 2010 in Dataset C contributed higher mean effects that were statistically significant. Given the rapid advancement of technology, it is encouraging to find increased effects among more recent research.

Results also indicate that CAI may be effective in both rural and urban areas, which increases the generalizability of these findings to similar regions of the country.

Participant characteristics. Compared with national averages, there was a higher proportion of diverse student populations as measured by race, socioeconomic status, and ELL status among the participants in this meta-analysis (NCES, 2012). This diversity and the large proportion of students reportedly at-risk in reading enhances the generalizability of these results to school populations with similar demographics. Furthermore, the finding of higher effects in urban areas is encouraging, as urban areas often report large achievement gaps between students in poverty and students without socioeconomic disadvantage (Walsh et al., 2014). The reading at-risk status of participants did not appear to affect outcomes, providing preliminary evidence that CAI may effectively improve the skills of all students regardless of at-risk status. Furthermore, the large proportion of participants with at-risk status contributing to mean effects in this meta-analysis indicates that CAI can effectively enhance the reading outcomes of at-risk students. This interpretation is tentative, as information on at-risk status in reading was not provided for more than half of all effects across datasets.

Pretest effect sizes were computed to investigate the impact of pretest differences on outcomes. In Datasets A and C, treatment groups with higher pretest scores than control groups tended to have higher effect sizes at posttest. In these datasets, a moderate pretest effect size (g ~ +0.50) was predicted to increase posttest effects by about 0.20 (Dataset A) and 0.34 (Dataset B). These findings suggest bias toward treatment groups with academic advantage at pretest compared to less advantaged control groups;

however, further analyses of pretest effects revealed a range from approximately -0.20 to +0.90, with unweighted means found to be around -0.15 for both datasets. These findings suggest that most participants in treatment groups did not have a pretest advantage. Many primary studies included in this analysis adjusted for pretest differences in the interpretations of their results; however, unadjusted means were used when provided, and results from adjusted versus unadjusted means were similar.

In Dataset B, gain scores were reported, providing an estimate of effect with pre-test differences already taken into account; however, the change-score metric has been shown to overestimate effects sizes and their variance (Morris & DeShon, 2002; Ray & Shaddish, 1996). In order for future meta-analyses of CAI in reading to be consistent, researchers should consider including enough information in reports to allow for meta-analysts to compute a common effect size metric across types of research design.

Intervention characteristics. Of the 101 independent effects, interventions most often occurred in a separate room (k = 47), lasted up to 30 min per session (k = 63), occurred over 3-5 sessions per week (k = 55), and lasted up to 50 hours in duration (k = 76). Licensed teachers led CAI most often (k = 49), though 26 independent effects came from studies not specifying the instructor.

Given findings that mixed or unspecified intervention location and literacy components contributed to higher effects across two datasets, the benefit of conducting interventions within core literacy in the classroom or supplemental to and separate from core literacy is inconclusive. The impact of total duration was mixed between datasets, but results suggest that interventions meeting less than five times per week, up to 30 min, and lasting less than 50 hours total yielded higher effects than those meeting five times per week for more than 30 min each session and/or lasting longer than 50 hours in total duration. Effects from studies lasting more than 50 hours total were often extracted from research on Fast ForWord, a program that led to mixed effects in this meta-analysis. The recommendation for implementation of Fast ForWord is a total duration of 40-60 hours, meeting five times per week, and in sessions lasting 30-100 minutes (U.S. Dept. of Education, 2013). Previous meta-analyses of reading interventions have noted that shorter interventions typically demonstrate higher effects (Elbaum, Vaughn, Hughes, & Moody, 2000; Wanzek et al., 2006).

Mean effects from CAI led by licensed teachers were positive and statistically significant in Datasets A and B, indicating that licensed teachers can implement CAI with positive results. In Datasets B and C, the moderate mean effects from CAI and teacher-led instruction provided by researchers was much higher than those provided by licensed teachers (g = +0.55 and +0.47, respectively). This finding is confounded by the fact that fidelity of implementation was either not assessed or not reported in about 80% of studies; thus, it is unclear what led to higher effects for researchers versus licensed teachers in those datasets. It is encouraging, however, that CAI led by licensed teachers can result in positive outcomes, enhancing the generalizability of these findings to existing classrooms.

CAI characteristics. Among 101 independent effects, commercially available software was most often described (k = 89) and all areas of reading were addressed by the programs relatively evenly. Results from the categorical moderator and meta-regression analyses were mixed. Commercially available programs yielded higher effects in Dataset A and lower effects in Dataset C.

Most programs were linear in nature (*k* = 65), meaning that students progressed through the program by completing sections sequentially, rather than the program adapting to student input by changing the sequence or difficulty level. The structure had little impact on effects in Datasets B and C, but the mean effect of adaptive programs was larger and statistically significant in Dataset A. This finding suggests that programs adapting to student input by changing the level of difficulty, sequence of activities, amount of scaffolding, and so forth may contribute to greater outcomes in reading. This finding is logical, as an adaptive program has the sophistication to provide for a more individualized, explicit, instructional experience for students, potentially enhancing feedback, providing guided activity, and increasing opportunities for students to respond correctly during sessions (Bishop & Santoro, 2006; Hall, 2002).

Across all datasets, the program components most often described were corrective feedback (*k* = 73), pacing (*k* = 59), and record keeping (*k* = 75). Analyses of program components are as difficult to interpret as areas of reading, since the inclusion of one component was not necessarily at the exclusion of another component. Preliminary evidence suggests that record keeping and guided activity may enhance reading outcomes; however, in this report those coded variables may or may not have included additional components. The number of components programs included did not play a consistent role in either explaining variation or enhancing effect sizes in one category over another.

Meta-regression across characteristics. Meta-regression analyses across study, participant, intervention, and CAI characteristics were conducted to further examine the impact of multiple methodological characteristics on mean effects. In Dataset A, a model including pretest effect size, experimental design, and teacher type accounted for almost 30% of the variation in effects. In this model, studies using a randomized control trial design were more likely to yield lower effects when pretest effects were high and licensed teachers or researchers implemented the treatment. In Dataset B, a model including the intervention characteristics of intervention location, literacy component, and total duration resulted in the greatest explained variation in effects (R^2 = 15.5%) and predicted higher effects for studies with mixed or unspecified literacy components, mixed or separate locations, and with increased total duration. In Dataset C, the location of the study, the intervention location, and the effect size at pretest explained 100% of the variation in effects, and predicted higher effects for studies conducted in urban areas, in mixed locations in the school, and with higher pretest effect sizes.

Though there is relative inconsistency in the results of the categorical moderator and meta-regression analyses between datasets, some areas of consistency exist. For example, pretest effect size had a large impact on two of the three datasets. Diverse populations with high proportions of academically and economically disadvantaged students in urban areas can achieve enhanced reading outcomes following the use of CAI. Intervention characteristics helped to explain a great deal of variability in effects in all three datasets, and provide some indication of adequate timing to receive benefit from CAI in reading. Furthermore, the presence of licensed teachers yielded positive, statistically significant mean effects in two of three datasets, enhancing the practical implications of using CAI to improve outcomes in reading.

What is the overall quality of the research according to the quality indicators proposed by Gersten et al. (2005)? The 61 studies meeting criteria for this review met about half of the quality indicators, on average. There were 42 published journal articles, 12 dissertations, and seven technical reports. When averaged across type of publication, there was little difference between the percentage of indicators met for journals, dissertations, or technical reports. In fact, technical reports outshined other publication types in terms of reporting effect sizes, timing measures appropriately, assessing fidelity, and providing evidence of comparability between groups and interventionists. Of the seven technical reports, two were published by the U.S. Department of Education, three by independent school districts, and one by the commercial producer of the software. Perhaps the high stakes of purchasing a program at the district level and/or establishing an evidence base for a commercially available program led researchers of these technical reports to adhere to quality research standards.

Across datasets, most studies met between 26%-50% of the quality indicators; the mean effect of studies in this category was statistically significant and similar to the summary effects in Datasets A and B. In Dataset C, most effects came from studies meeting 51%-75% of the quality indicators, suggesting somewhat higher quality among this set of studies. Very few studies (n = 13) met 75% or more of the quality indicators. In the meta-regression and categorical moderator analyses, the number of indicators met (i.e., the quality score) did not emerge as a factor affecting mean effects.

Relative strengths of the research included providing adequate descriptions of the intervention, timing measures appropriately, and providing an appropriate data analysis. Several indicators of high-quality research were less prevalent among this set of studies including providing evidence of comparable groups and interventionists, reporting attrition, assessing fidelity, and assessing outcomes with multiple measures.

Among the meta-regression models of quality indicators, the use of multiple measures, comparable interventionists and reporting of attrition explained a portion of the variation across datasets. Studies reporting multiple measures typically yielded higher effect sizes. Providing evidence of comparability between groups or interventionists did not always lead to higher effects, suggesting potential bias favoring studies not meeting this indicator. This finding is not surprising, as studies with larger pretest effect sizes favoring treatment were given a code of zero on indicators of group comparability. Suggestions for enhancing future research quality are discussed in the future research directions section.

Comparison with Previous Research on CAI in Reading

A review of the literature (see Chapter 2) revealed four narrative reviews and nine meta-analyses including outcomes of CAI in reading examining a total of 224 studies (see Table 2, p. 40). The number of studies reviewed in each synthesis ranged from 5-82, with a median of 18. Situated within this research, the meta-analysis reported in this dissertation examined more studies than all but one previous review (Cheung & Slavin, 2012b; $n = 82$). Reported mean effects of CAI in previous syntheses ranged from -0.03 to +0.49, with a median effect of +0.13. Only one review found evidence of negative effects of CAI (Strong et al., 2011). Only 14 of the studies included in this review overlapped with the 71 studies conducted since 2000 that were reported in previous syntheses. This was not surprising, as the focus of many previous syntheses was on the broad use of technology in reading education, and included studies examining assistive technology and/or educational technology that did not meet eligibility criteria for this meta-analysis. Additionally, many previous meta-analyses limited their search to peer-reviewed journal articles, thus excluding the dissertations and technical reports included in this meta-analysis.

Several previous meta-analyses examined moderators to explain differences among mean effect sizes. Similar to this meta-analysis, Blok et al. (2002) found that treatment groups displaying advantage at pretest tended to attain higher effects at post-test. They also found that studies conducted in English yielded a higher, moderate mean effect (g = +0.50) than studies conducted in other languages. Moran et al. (2008) found that studies with more than 30 participants using researcher-designed measures yielded higher effects. They also found that commercially-available software yielded a smaller effect than the overall mean; however, this effect was still positive and statistically significant (g = +0.28). Mean effects of commercially-available software in Datasets A and B in this analysis also reflected those findings, with effect sizes of +0.21 and +0.13 respectively. Evidence from Moran et al.'s (2008) study suggested that CAI may be more effective for students in general education than special education. The results of this meta-analysis suggest that the percentage of participants at-risk in reading or with learning disabilities did not impact reading outcomes. Also similar to the results of this meta-analysis, Moran et al. (2008) did not find differences among the areas of reading assessed in included studies.

Cheung and Slavin (2011, 2012a, 2012b) conducted several meta-analyses of CAI and situated it within analyses of other instructional delivery types. The mean effects for CAI were similar to those found in this meta-analysis (range = +0.09-0.18), and moderator analyses revealed that large samples using a randomized control trial design resulted in lower mean effects. Published reports had higher effect sizes than unpublished reports. In this meta-analysis, results from moderators of sample size, experimental design and publication type varied across datasets.

Cheung and Slavin's (2011, 2012a, 2012b) work provides an avenue to examine CAI in the context of instructional delivery. Their findings indicate that the effects of CAI in reading are similar when compared to one-on-one tutoring from volunteers (g = +0.16) and lower than classroom-based comprehensive programs such as Success For All (g = +0.55; Cheung & Slavin, 2011). Results of comprehensive programs utilizing technology but primarily instructed by a teacher (e.g., Read 180), were mixed, ranging from +0.04 (Cheung & Slavin, 2012a) to +0.28 (Cheung & Slavin, 2012b).

This meta-analysis focused on investigating the effects of instruction in reading primarily delivered through a computer. The work of Cheung and Slavin (2011, 2012a, 2012b) provides preliminary evidence that more comprehensive, teacher-led programs might produce larger effects than instruction delivered through the computer. Meta-analyses of more traditional methods of intervention delivery, such as one-to-one tutoring with a licensed teacher, have yielded effect sizes of larger magnitude than those found in the current or previous meta-analyses, with effect sizes ranging from +0.41-0.95 (Edmonds et al., 2009; Flynn, Zheng, Swanson, 2012; Scammacca et al., 2007; Swanson, 1999). While this finding might suggest that instruction delivered through computers is not as effective as instruction delivered by a teacher, this comparison should be interpreted with caution. In order to truly compare different types of instructional delivery, researchers should consider including a comparison group receiving teacher-led intervention targeting the same area(s) of reading and providing similar strategies and content, much like the studies conducted within Dataset C. Meta-analyses of instruction in reading including both CAI and teacher-led instruction (e.g., Cheung and Slavin, 2011) should analyze the different characteristics of interventions that may lead to higher effects in teacher-led instruction vs. instruction delivered via computer. The results of Dataset C indicate little difference between groups receiving CAI or teacher-led instruction; further investigation of potential differences is warranted before definitive conclusions can be drawn.

Practical Implications

The findings of positive, statistically significant effects when comparing CAI to no treatment or non-reading CAI are encouraging for educators planning to use CAI to improve the reading outcomes of students in preschool through high school. The finding of no difference between groups receiving CAI and groups receiving the same targeted instruction delivered by a teacher was also encouraging, indicating that CAI may be as effective and potentially more efficient than teacher-led, small-group intervention in reading.

Given a large proportion of diversity in race, academic disadvantage, and socio-economic status in the participant samples of this meta-analysis, educators working in diverse, urban areas can use CAI as a viable instructional delivery tool for students at all school levels struggling in reading. Results suggest that licensed teachers and researchers can obtain positive effects when implementing CAI with students. Given the nature of CAI as an instructional delivery tool, this finding is encouraging. Since inclusion criteria limited this meta-analysis to studies investigating CAI primarily delivered via a computer or handheld device, it would be expected that the person implementing the intervention (teacher vs. researcher) would not have a great impact on outcomes. Considering the individualized nature of CAI, it is possible that large groups of students monitored by one teacher could receive one-on-one instruction and practice through the computer, perhaps in the school's computer lab. If other school staff or volunteers could monitor the use of CAI, this could free up licensed teachers to conduct more intensive, face-to-face interventions with students who have significant needs in the area of reading; however, other school staff or volunteers were only represented in 33 effects, and outcomes were inconclusive.

Tables C1, C2, and D1 (pp.178-190) provide information on CAI characteristics categorized by program name. This information might be beneficial to educators choosing CAI in reading as an instructional delivery tool. The results of this meta-analysis suggest that adaptive programs providing guided activity and record keeping may yield higher effects, though these findings are preliminary and not at the exclusion of other program components. The impact of an increased number of program components was inconclusive in this analysis. Thus, educators choosing CAI should not necessarily look for programs that have numerous bells and whistles, but should look for evidence of instructional design consistent with suggestions by Moreno and Mayer (2007) and Santoro & Bishop (2010) while also aligning program selection with students' specific reading needs.

Considering these results in the context of cognitive learning theory (Chandler & Sweller, 1991; Kreskey, 2012; Merrienboer & Sweller, 2005; Sweller et al., 1998) and the accompanying cognitive theory of multimedia learning (Bishop & Santoro, 2006; Moreno & Mayer, 2007; Smith & Okolo, 2006), educators should take precautions when choosing programs with components that may strain cognitive resources by presenting extraneous load. For example, a program with a lot of colors and animations that do not contribute to the students' learning could potentially result in decreased cognitive capacity to learn and apply new material (Moreno & Mayer, 2007). Educators have to weigh the importance of limiting extraneous load in the programs they choose with student motivation, a component of instructional design not often described in CAI research.

One moderator that consistently explained differences in effect sizes was the quality indicator of multiple measures. Mean effects were consistently higher for groups of studies utilizing multiple measures than those using a single measure of effect. This is important for practice, as teachers should examine student progress in reading using both measures targeting the instructional focus as well as measures of broad reading. When examining the effectiveness of CAI in reading, multiple measures may help to capture intervention effectiveness or lack thereof, thus helping to inform instructional decisions through adequate data.

Although differential results across school levels and areas of reading were inconclusive, it was interesting to note that results for early reading measures (e.g., phonics and phonemic awareness) were higher in Dataset A, which had a higher proportion of younger students and results for comprehension were higher in Dataset B, which had a higher proportion of older students. Given the finding that measures aligned with the intervention tended to produce higher effects, educators seeking CAI in reading should consider the skill deficits, instructional level, and developmental level of the students who will be receiving instruction. If an elementary student struggles with basic reading skills, for example, a short-term intervention targeting comprehension is unlikely to demonstrate effects. The same may likely be true for a middle or high school student. Student data should drive instructional decisions, CAI aligned to student need should be selected, and multiple measures both aligned with the intervention and measuring broad reading skills should be used to measure student progress.

Limitations and Future Research Directions

In an effort to include as many studies as possible in this meta-analysis, dissertations and technical reports were included in addition to published journal articles. Analyses of publication bias indicated little difference between mean effects of each publication type and the fail-safe *N* ranged from 178 to 1,070, indicating many more studies would be required in each dataset to nullify results. Though evidence of publication bias was not present in this review, the exclusion criteria prevented the examination of some studies investigating CAI in reading. For example, studies using a large, pre-posttest design with no control group and those using single-case designs did not meet criteria for review in this meta-analysis. Future researchers should analyze the impact of these studies on reading outcomes following the use of CAI and compare the findings to those in this and previous meta-analyses of studies using group design.

Conclusions regarding differential effectiveness across school level and areas of reading were inconclusive. Though a variable assessing the alignment of the intervention with assessment was created to capture the effectiveness of more targeted interventions, many dependent measures assessed multiple areas of reading, thus convoluting interpretations. Future research investigating CAI targeted to student need and assessed with specific and broad measures should be conducted to investigate the effectiveness of CAI as an intervention delivery tool in a tiered model. This type of research may also help to parse out the effectiveness of CAI at different school levels, since CAI addressing comprehension skills, for example, may result in different outcomes for elementary versus secondary students.

Preliminary evidence points to specific instructional design components (i.e., guided activity and record keeping) that may strengthen outcomes. These characteristics should be further investigated through both primary and meta-analytic research, as the efficiency of treatment in a tiered model of intervention delivery must be considered when allocating resources. Primary research investigating the role of specific instructional design components in CAI would help in the determination of effective software design. If programs can be implemented for shorter durations and still exhibit positive effects, the components of those programs should be clearly described in order to further the evidence base of instructional design in CAI. Additionally, primary investigations of CAI led by volunteers or unlicensed school support staff would provide evidence that CAI can be used as a stand-alone tool regardless of the professional accreditation of the person monitoring the intervention.

The coding of research quality in this meta-analysis revealed deficiencies in the evidence-base of CAI in reading. A limitation of this coding procedure was the dichotomous nature of the coding manual. If a rubric was used for the coding process, similar to those used in other syntheses (e.g., Baker et al., 2009; Jitendra et al., 2011), more information regarding the contribution of research quality may have been gleaned. The findings of this analysis indicate that future research should address the comparability of groups prior to treatment, the comparability of interventionists across groups, attrition rates, fidelity of implementation, and the use of multiple measures to assess outcomes. In addition, researchers should consider reporting pretest scores, their standard deviations, and unadjusted posttest scores to allow for more robust comparisons. Had the effects from Dataset B been combined with those in Dataset A, perhaps more consistency in the results of the moderator analyses could have been attained.

There was great variation between large (>250) and small (<250) studies in this meta-analysis and the impact of sample size was mixed. The use of a random-effects model tends to even out the contribution of large studies with increased heterogeneity; thus, in Dataset A, effects were weighted relatively evenly. The precision of the effect size estimates in the large study conducted by Rasinski et al. (2011) resulted in larger weights assigned to the effects extracted from this study. Given the recommendation by the Institute of Educational Sciences to include at least 300 participants in randomized control trials (U.S. Dept. of Education, 2003), large studies such as Rasinski et al. (2011) with precise effect size estimates should be further investigated for characteristics leading to effects or lack thereof.

Another limitation of this review relates to the ever-changing nature of technology. With the rise of handheld devices in classrooms around the country, studies investigating touchscreen devices were either not located for review or failed to meet the inclusion criteria of this meta-analysis. This may reflect the lag between conducting and publishing research, and this finding highlights the need for current research of this instructional delivery type. As teachers more frequently rely on applications purporting to enhance reading outcomes, research investigating the effectiveness of these programs is essential. In addition, meta-analyses of current research should be conducted frequently, in order to capture the changing nature of technology as well as the impact of more sophisticated programs and ease of access on outcomes in reading.

Last, though evidence of motivation was included in the coding manual, the vast majority of studies did not assess student motivation, engagement, or attitudes toward the CAI being used. As technology becomes more and more integrated into the everyday lives of children, motivation and attitudes toward CAI should be assessed in every primary study. This would help educators and researchers determine not only which programs might better engage students, but also what types of components lead to increased motivation and engagement in reading CAI.

Conclusion

This study investigated the impact of computer-assisted instruction on the reading outcomes of students in preschool through high school. By including studies published since 2000, this study contributes to the evidence base regarding the effectiveness of CAI in reading. Through meta-analytic procedures, the impact of study and intervention characteristics could be examined, revealing information for educators to consider when choosing CAI to use with their students.

Overall, preschool through high school students receiving CAI experienced more progress in reading than their peers not receiving this type of instruction. The generalizability of the findings in this review are enhanced by the fact that diversity in race, socio-economic status, and risk for reading was represented in the overall sample, and studies using licensed teachers yielded positive effects. Schools are using technology on a daily basis and the technology they use is updated and refined at a staggering pace. Future, high-quality research should be conducted to continually evaluate changing technology and help to inform both software developers and educators regarding the most effective instructional methods to use in reading CAI.

When examining available interventions and weighing the costs and benefits of CAI versus traditional teacher-led intervention delivery, educators can look to the expansive research base for evidence that CAI is a viable option for enhancing students' reading skills. In an era of increasing focus on data- and evidence-based intervention delivery, CAI has the potential to supply a needed instructional resource providing targeted, individualized instruction to numerous students simultaneously. Through the use of computer-assisted instruction, efficient, effective, and individualized interventions can be delivered, providing all students with the opportunity to experience growth and success in the area of reading.

References

A supplemental reference list of studies included in the meta-analyses can be located in Appendix H.

Adams, G. L., & Engelmann, S. (1996). *Research on direct instruction: 25 years beyond DISTAR* Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED413575&site=ehost-live

Ae-Hwa Kim, Vaughn, S., Klingner, J. K., Woodruff, A. L., Reutebuch, C. K., & Kouzekanani, K. (2006). Improving the reading comprehension of middle school students with disabilities through computer-assisted collaborative strategic reading. *Remedial & Special Education*, *27*(4), 235-249. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=21639817&site=ehost-live

Annie E. Casey Foundation (2010). *Early warning! Why reading by the end of third grade matters.* (Special Report). Baltimore, MD. Retried from http://www.aecf.org/resources/early-warning-why-reading-by-the-end-of-third-grade-matters/

Atkinson, R. C. (1968). Computerized instruction and the learning process. *American Psychologist, 23*(4), 225-239. doi:10.1037/h0020791

Baker, S. K., Chard, D. J., Ketterlin-Geller, L. R., Apichatabutra, C., & Doabler, C. (2009). Teaching writing to at-risk students. The quality of evidence for self-regulated strategy development. *Exceptional Children*, *75*(3), 303-318.

Basham, J. D., Israel, M., Graden, J., Poth, R., & Winston, M. (2010). A comprehensive approach to RTI: Embedding universal design for learning and technology. *Learning Disability Quarterly, 33*(4), 243-255. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=55486635&site=ehost-live

Becker, H. J. (1992). Computer-based integrated learning systems in the elementary and middle grades: A critical review and synthesis of evaluation reports. Journal of Educational Computing Research, 8(1), 1–41.

Berkeley, S., Bender, W. N., Peaster, L. G., & Saunders, L. (2009). Implementation of response to intervention: A snapshot of progress. *Journal of Learning Disabilities, 42*(1), 85-95. doi: 10.1177/0022219408326214

Bishop, M. J., & Santoro, L. E. (2006). Evaluating beginning reading software for at-risk learners. *Psychology in the Schools, 43*(1), 57-70. doi:10.1002/pits.20129

Blok, H., Oostdam, R., Otter, M. E., & Overmaat, M. (2002). Computer-assisted instruction in support of beginning reading instruction: A review. *Review of Educational Research, 72*(1), 101-30.

Boone, R., & Higgins, K. (2007). The role of instructional design in assistive technology research and development. *Reading Research Quarterly*, *42*(1), 135-140.

Borenstein L.V., Hedges, J.P., Higgins, T., Rothstein, H.R. (2009) *Introduction to Meta-Analysis.* John Wiley & Sons, Ltd. ISBN: 978-0-470-05724-7

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.),*The Handbook of Research Synthesis and Meta-Analysis* (2nd ed.). New York: Russel Sage Foundation.

Brooks, G., Miles, J. N. V., Torgerson, C. J., & Torgerson, D. J. (2006). Is an intervention using computer software effective in literacy learning? A randomised controlled trial. *Educational Studies (03055698), 32*(2), 133-143. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=21007439&site=ehost-live

Burns, M. K., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analytic review of responsiveness-to-intervention research: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment, 23*(4), 381-394. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ806494&site=ehost-live; http://dx.doi.org/10.1177/073428290502300406

Campuzano, L., Dynarski, M., Agodini, R., Rall, K., & Institute of Education Sciences (ED), National Center for Education Evaluation and, Regional Assistance. (2009). *Effectiveness of reading and mathematics software products: Findings from two student cohorts. NCEE 2009-4041.*National Center for Education Evaluation and Regional Assistance. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED504657&site=ehost-live

Chall, J. S. (1983). Stages of reading development. New York: McGraw-Hill.

Chall, J. S. & Jacobs, V. A. (2003). *Poor children's fourth-grade slump*. Retrieved from http://www.colorincolorado.org/article/13995/

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293-332. Retrieved fromhttp://www.jstor.org/stable/3233596

Chera, P., & Wood, C. (2003). Animated multimedia 'talking books' can promote phonological awareness in children beginning to read. *Learning and Instruction, 13*(1), 33-52. doi: 10.1016/S0959-4752(01)00035-4

Cheung, A., Slavin, R.E. (2012). *The Effectiveness of Educational Technology Applications for Enhancing Reading Achievement in K-12 Classrooms: A Meta-Analysis*. Baltimore, MD: Johns Hopkins University, Center for Research and Reform in Education.

Cheung, A., & Slavin, R. (2012a). Effects of educational technology applications on reading outcomes for struggling readers: A best evidence synthesis. *Best Evidence Encyclopedia*, Retrieved from http://www.bestevidence.org/

Cheung, A., & Slavin, R. (2012b). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review, 7*(3), 198-215.

Christmann, E. P., & Badgett, J. L. (2003). A Meta-analytic comparison of the effects of computer-assisted instruction on elementary students' academic achievement. *Information Technology in Childhood Education*, 91-104.

Clay, M. (1979). The early detection of reading difficulties. (3rd ed). Auckland, New Zealand: Heinemann.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cole, J. M., & Hilliard, V. R. (2006). The effects of web-based reading curriculum on children's reading performance and motivation. *Journal of Educational Computing Research*, *34*(4), 353-380. Retrieved fromhttp://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ757162&site=ehost-live; http://baywood.metapress.com/openurl.asp?genre=article&id=doi:10.2190/H43W-1N3U-027J-07V5

Comprehensive Meta-Analysis (Version 2) [Computer software]. Englewood, NJ: Biostat.

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). Research synthesis as a scientific process. In H. Cooper, L. Hedges & J. Valentine (Eds.), *Handbook of Research Synthesis and Meta-Analysis* (2nd ed., pp. 3-16). New York: Russell Sage Foundation.

Denson, N., & Seltzer, M. (2011). Meta-analysis in higher education: An illustrative example using hierarchical linear modeling. *Research In Higher Education*, *52*(3), 215-244. doi:10.1007/s11162-010-9196-x

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455-463. Retrieved fromhttp://www.jstor.org/stable/2676988

Dynarski, M., Agodini, R., Heaviside, S. N., Timothy, Carey, N., Campuzano, L., Means, B. et al., IES (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort. Report to congress.* National Center for Education Evaluation and Regional Assistance. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED496015&site=ehost-live

Edmonds, M. S., Vaughn, S., Wexler, J., Reutebuch, C., Cable, A., Tackett, K. K., et al. (2009). A synthesis of reading interventions and effects on reading comprehension outcomes for older struggling readers. *Review of Educational Research, 79*(1), 262-300. Retrieved fromhttp://www.jstor.org/stable/40071166

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. Scientific Studies of Reading, 9, 167-188.

Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S. W. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology, 92, 605–619. doi:10.1037/0022-0663.92.4.605*

Fletcher, J. D., & Atkinson, R. C. (1972). *Evaluation of the Stanford CAI program in initial reading* Journal of Educational Psychology. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ068074&site=ehost-live

Fletcher, J. M., & Vaughn, S. (2009). Response to intervention: Preventing and remediating academic difficulties. *Child Development Perspectives, 3*(1), 30-37. doi:10.1111/j.1750-8606.2008.00072.x

Fletcher-Finn, C., & Gravatt, B. (1995). The efficacy of computer-assisted instruction (CAI): A meta-analysis. *Journal of Educational Computing Research, 12*(3), 219–241.

Flynn, L. J., Zheng, X., & Swanson, H. (2012). Instructing struggling older readers: A selective meta-analysis of intervention research. *Learning Disabilities Research & Practice*, *27*(1), 21-32. doi:10.1111/j.1540-5826.2011.00347.x

Foorman, B. R. (2007). Primary prevention in classroom reading instruction. *Teaching Exceptional Children, 39*(5), 24-30. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=25204850&site=ehost-live

Foorman, B. R., & Torgesen, J. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. *Learning Disabilities Research & Practice, 16*(4), 203-212. doi:10.1111/0938-8982.00020

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2012). Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children, 78*(3), 263-279. Retrieved from http://search.ebscohost.com.ezp1.lib.umn.edu/login.aspx?direct=true&db=aph&AN=72104449&site=ehost-live

Gale, Deanne (2006) The Effect of Computer-Delivered Phonological Awareness Training on the Early Literacy Skills of Students Identified as At-Risk for Reading Failure (Doctoral dissertation). *Graduate School Theses and Dissertations.* Retrieved from http://scholarcommons.usf.edu/etd/3860

Gatti, G. G. & Miller, J. (2011). *SuccessMaker: Pearson SuccessMaker reading efficacy study. 2010-11 Final Report*. Pittsburgh, PA: Gatti Evaluation Inc. Retrieved from: https://www.pearsoned.com/wp-content/uploads/sm-reading-rct-report1.pdf

Gersten, R., Compton, D., Connor, C.M., Dimino, J., Santoro, L., Linan-Thompson, S., and Tilly, W.D. (2008). *Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading in the primary grades. A practice guide.* (NCEE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/.

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*(2), 149-164. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=15457832&site=ehost-live\

Gray, L., Thomas, N., and Lewis, L. (2010). *Teachers' Use of Educational Technology in U.S. Public Schools: 2009* (NCES 2010-040). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Greenhouse, J. B. & Iyengar, S. (2009). Sensitivity analysis and diagnostics. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The Handbook of Research Synthesis and Meta-Analysis* (2nd ed.). New York: Russel Sage Foundation.

Greenlee, N. B. (2001). An analysis of the effectiveness of the "academy of reading" training on the reading achievement of elementary students in a suburban, Midwestern school district. (Doctoral dissertation). *ProQuest Dissertations and Theses*, Retrieved fromhttp://search.proquest.com.ezp1.lib.umn.edu/docview/230734556?accountid=14586

Grigg, W. S., Daane, M. C., Jin, Y., Campbell, J. R., Educational, T. S., & National Center for, E. S. (2003). *The nation's report card: Reading, 2002.* Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED471794&site=ehost-live

Hall, T. (2002). *Explicit instruction: Effective classroom practices report.* Wakefield, MA: National Center on Accessing the General Curriculum. Retrieved fromhttp://www.mdecgateway.org/olms/data/resource/7595/ExpInstrucCAST.doc

Hall, T. E., Hughes, C. A., & Filbert, M. (2000). Computer assisted instruction in reading for students with learning disabilities: A research synthesis. *Education & Treatment of Children, 23*(2), 173.

Hansen, E. E., Llosa, L., & Slayton, J. (2004). *Evaluation of the Waterford Early Reading Program as a supplementary program in the Los Angeles Unified School District: 2002-03.* (Publication No. 177). Retrieved from http://notebook.lausd.net/portal/page?_pageid=33,103101&_dad=ptl

Hasselbring, T. S., & Goin, L. I. (2004). Literacy instruction for older struggling readers: What is the role of technology? *Reading and Writing Quarterly, 20*(2), 123-144.

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

Hernandez, D. J. (2011). *Double jeopardy: How third-grade reading skills and poverty influence high school graduation.* Retrieved from The Annie E. Casey Foundation website: http://www.aecf.org/resources/double-jeopardy/

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta- analyses. *BMJ (Clinical Research Ed.), 327*(7414), 557.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539-1558. doi:10.1002/sim.1186

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. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=a2h&AN=15457833&site=ehost-live

Horton, S. V., Lovitt, T. C., & Givens, A. (1988). A computer-based vocabulary program for three categories of student. *British Journal of Educational Technology*, *19*(2), 131-143.

Huedo-Medina, T., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: *Q* statistic or I^2 index? *Psychological Methods*, *11*(2), 193-206. doi:10.1037/1082-989X.11.2.193

Huffstetter, M., King, J. R., Onwuegbuzie, A. J., Schneider, J. J., & Powell-Smith, K. (2010). Effects of a computer-based early reading program on the early reading and oral language skills of at-risk preschool children. *Journal of Education for Students Placed at Risk (JESPAR)*, *15*(4), 279-298. doi:10.1080/10824669.2010.532415

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.. Newbury Park: Sage.

Individuals with Disabilities Education Improvement Act of 2004, Pub. L. No. 108-446 (2004) Retrieved from http://idea.ed.gov/.

Jenkins, J. R., Schiller, E., Blackorby, J., Thayer, S., & Tilly, W. (2013). Responsiveness to Intervention in Reading: Architecture and Practices. *Learning Disability Quarterly*, *36*(1), 36-46. doi:10.1177/0731948712464963

Jimerson, S. R., Burns, M. K., & VanDerHeyden, A. M. (2007). *Handbook of response to intervention*: *The science and practice of assessment and intervention*. New York: Springer.

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*(2), 135-159. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=57845590&site=ehost-live

Jitendra, A. K., Edwards, L. L., Sacks, G., & Jacobson, L. A. (2004). What research says about vocabulary instruction for students with learning disabilities. *Exceptional Children*, *70*(3), 299. Retrieved fromhttp://search.ebscohost.com.ezp2.lib.umn.edu/login.aspx?direct=true&db=eric&AN=EJ696073&site=ehost-live

Jitendra, A. K., Petersen-Brown, S., Lein, A. E., Zaslofsky, A. F., Kunkel, A. K., Jung, P., et al. (2015). Teaching mathematical word problem solving: The quality of evidence for strategy instruction priming the problem structure. *Journal of Learning Disabilities*, *48*(1), 51-72. doi:10.1177/0022219413487408

Jones, J. D., Staats, W. D., Bowling, N., Bickel, R. D., Cunningham, M. L., & Cadle, C. (2004). An evaluation of the merit reading software program in the Calhoun county (WV) Middle/High school. *Journal of Research on Technology in Education, 37*(2), 177-195. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=15306380&site=ehost-live

Kamil, M. L., Borman, G. D., Dole, J., Kral, C. C., Salinger, T., and Torgesen, J. (2008). *Improving adolescent literacy: Effective classroom and intervention practices: A Practice Guide* (NCEE #2008-4027). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc.

Kennedy, M. J., & Deshler, D. D. (2010). Literacy instruction, technology, and students with learning disabilities: Research we have, research we need. *Learning Disability Quarterly, 33*(4), 289-298. Retrieved fromhttp://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=55486638&site=ehost-live

Kim, A., Vaughn, S., Klingner, J. K., Woodruff, A. L., Reutebuch, C. K., & Kouzekanani, K. (2006). Improving the reading comprehension of middle school students with disabilities through computer-assisted collaborative strategic reading. *Remedial & Special Education*, *27*(4), 235-249.

King Thorius, K. A., Maxcy, B. D., Macey, E., & Cox, A. (2014). A critical practice analysis of response to intervention appropriation in an urban school. *Remedial and Special Education*, *35*(5), 287-299. doi:10.1177/0741932514522100

Koury, K. A. (1996). The impact of preteaching science content vocabulary using integrated media for knowledge acquisition in a collaborative classroom. *Journal of Computing in Childhood Education*. 7(3/4), 179-197.

Kovaleski, J. F. (2007). Response to intervention: Considerations for research and systems change. *School Psychology Review, 36*(4), 638.

Kreskey, D. D. (2012). Headsprout early reading for students at risk for reading failure. (Ph.D., Georgia State University). *ProQuest Dissertations and Theses,* Retrieved from http://search.proquest.com.ezp2.lib.umn.edu/docview/1037809369?accountid=14586

Kulik, C. C., & Kulik, J. A. (1991). Effectiveness of computer-based instruction: An updated analysis. *Computers in Human Behavior,7*(1–2), 75-94.

Kulik, J. A. (1994). Meta-analytic studies of findings on computer-based instruction. In E. Baker & H. O'Neil (Eds.), *Technology Assessment in Education and Training* (pp. 9-33). New Jersey: Lawrence Erlbaum Associates

Kulik, J. A. (2003). *Effects of using instructional technology in elementary and secondary schools: What controlled evaluation studies say* (SRI International Report). Retrieved from: http://citeseerx.ist.psu.edu/viewdoc/download?rep=rep1&type=pdf&doi=10.1.1.207.3105

Kunkel, A. K. (2014). *The effects of computer-assisted instruction in reading: Review of research syntheses.* Manuscript in preparation.

Lembke, E. S., McMaster, K. L., & Stecker, P. M. (2010). The prevention science of reading research within a response-to-intervention model. *Psychology in the Schools, 47*(1), 22-35. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ871962&site=ehost-live; http://www.blackwell-synergy.com/doi/abs/10.1002/pits.20449

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publishing.

Littleton, K., Wood, C., & Chera, P. (2006). Interactions with talking books: Phonological awareness affects boys' use of talking books. *Journal of Computer Assisted Learning, 22*(5), 382-390. doi:10.1111/j.1365-2729.2006.00183.x

MacArthur, C. A., Ferretti, R. P., Okolo, C. M., & Cavalier, A. R. (2001). Technology applications for students with literacy problems: A critical review. *The Elementary School Journal, 101*(3), 273-301. Retrieved from http://www.jstor.org/stable/1002248

Macaruso, P., Hook, P. E., & McCabe, R. (2006). The efficacy of computer-based supplementary phonics programs for advancing reading skills in at-risk elementary students. *Journal of Research in Reading, 29*(2), 162-172.

Macaruso, P., & Rodman, A. (2009). Benefits of computer-assisted instruction for struggling readers in middle school. *European Journal of Special Needs Education, 24*(1), 103-113.

Macaruso, P., & Walker, A. (2008). The efficacy of computer-assisted instruction for advancing literacy skills in kindergarten children. *Reading Psychology, 29*(3), 266-287. doi:10.1080/02702710801982019

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 1-16). New York: Cambridge University Press

McKenna, M. C., & Walpole, S. (2007). Assistive technology in the reading clinic: Its emerging potential. *Reading Research Quarterly*, *42*(1), 140-145.

Mitchell, M. J., & Fox, B. J. (2001). The effects of computer software for developing phonological awareness in low-progress readers. *Reading Research and Instruction*, *40*(4), 315-332. doi: 10.1080/19388070109558353

Moran, J., Ferdig, R., Pearson, P. D., Wardrop, J., & Blomeyer, R. L. (2008). Technology and reading performance in the middle-school grades: A meta-analysis with recommendations for policy and practice. *Journal of Literacy Research*, *40*, 6-58.

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review, 19*(3), 309-326. doi:10.1007/s10648-007-9047-2

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*(1), 105-125. doi:10.1037/1082-989X.7.1.105

National Center for Education Statistics (2011). The Nation's Report Card: Reading 2011(NCES 2012–457). Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

National Institute of Child Health and Human Development. (2000). *Report of the national reading panel. Teaching children to read: An evidence-based assessment* *of the scientific research literature on reading and its implications for reading instruction*. (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

Nicolson, R. I., Fawcett, A. J., & Nicolson, M. K. (2000). Evaluation of a computer-based reading intervention in infant and junior schools. *Journal of Research in Reading, 23*(2), 194. Retrieved fromhttp://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=10452916&site=ehost-live

Ouyang, J. (1993). *Meta-analysis: CAI at the level of elementary education*. Paper presented at the World Conference on Education Multimedia and Hypermedia. (retrieved from Cheung & Slavin, 2012a)

Pinkard, N. (2001). Rappin' reader and say say oh playmate: Using children's childhood songs as literacy scaffolds in computer-based learning environments. *Journal of Educational Computing Research*, *25*(1), 17-34.

Potter, M. L., & Wamre, H. M. (1990). Curriculum-based measurement and developmental reading models: opportunities for cross-validation. *Exceptional Children*, *57*(1), 16+.

R Development Core Team (2008). [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rasinski, T., Samuels, S. J., Hiebert, E., Petscher, Y., & Feller, K. (2011). The relationship between a silent reading fluency instructional protocol on students' reading comprehension and achievement in an urban school setting. *Reading Psychology,32*(1), 75-97.

Raudenbush, S. W., Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Ray, J. W., & Shadish, W. R. (1996). How interchangeable are different estimators of effect size? *Journal of Consulting and Clinical Psychology, 64*(6), 1316-1325. doi:10.1037/0022-006X.64.6.1316

Rouse, C.E., & Krueger, A.B. (2004). Putting computerized instruction to the test: A randomized evaluation of a 'scientifically based' reading program. *Economics of Education Review*, 23, 323–338.

Rupley, W. H., Blair, T. R., & Nichols, W. D. (2009). Effective reading instruction for struggling readers: The role of Direct/Explicit teaching. *Reading & Writing Quarterly, 25*(2), 125-138. doi:10.1080/10573560802683523

Santoro, L., & Bishop, M. J. (2010). Selecting software with caution: An empirical evaluation of popular beginning reading software for children with early literacy difficulties. *Computers in The Schools*, *27*(2), 99-120.

Scammacca, N., Roberts, G., Vaughn. S., Edmonds, M., Wexler, J., Reutebuch, C. K., & Torgesen, J. K. (2007). *Interventions for adolescent struggling readers: A meta-analysis with implications for practice*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Shadish, W. R., & Haddock, C. K. (2009). Combining estimates of effect size. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The Handbook of Research Synthesis and Meta-Analysis* (2nd ed.). New York: Russel Sage Foundation.

Shanahan, T., & North Central Regional, E. L. (2005). *The national reading panel report. practical advice for teachers.* Learning Point Associates / North Central Regional Educational Laboratory (NCREL). Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED489535&site=ehost-live

Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher, 15*(9), 5-11. Retrieved from http://www.jstor.org/stable/1174711

Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review, 6*(1), 1-26. Retrieved from http://search.ebscohost.com.ezp2.lib.umn.edu/login.aspx?direct=true&db=eric&AN=EJ916476&site=ehost-live;http://dx.doi.org.ezp2.lib.umn.edu/10.1016/j.edurev.2010.07.002

Smith, S. J., & Okolo, C. (2010). Response to intervention and evidence-based practices: Where does technology fit? *Learning Disability Quarterly, 33*(4), 257-272. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=55486636&site=ehost-live

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, D.C.: National Academy Press. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=991&site=ehost-live

Soe, K., Koki, S., & Chang, J. Pacific Resources for Education and Learning, (2000). *Effect of computer-assisted instruction (CAI) on reading achievement: A meta-analysis.* Retrieved from http://www.cs.cmu.edu/afs/cs.cmu.edu/project/listen/dmg/Listen/Proposals/IERI%202003/Effect%20of%20Computer-Assisted%20Instruction%20(CAI)%20on%20Reading%20Achievement%20A%20Meta-Analysis.htm

SPSS (Version 22) [Computer software]. Armonk, NY: IBM Corp.

Stein, M., Johnson, B., & Gutlohn, L. (1999). Analyzing beginning reading programs: The relationship between decoding instruction and text. *Remedial and Special Education, 20*(5), 275-287. doi:10.1177/074193259902000503

Stetter, M. E., & Hughes, M. T. (2010). Computer-assisted instruction to enhance the reading comprehension of struggling readers: A review of the literature. *Journal of Special Education Technology, 25*(4), 1-16. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1002175&site=ehost-live; http://www.tamcec.org/jset-index/computer-assisted-instruction-to-enhance-the-reading-comprehension-of-struggling-readers-a-review-of-the-literature-3/

Strong, G. K., Torgerson, C. J., Torgerson, D., & Hulme, C. (2011). A systematic meta-analytic review of evidence for the effectiveness of the "fast ForWord" language intervention program. *Journal of Child Psychology and Psychiatry, 52*(3), 224-235. Retrieved from http://search.ebscohost.com.ezp2.lib.umn.edu/login.aspx?direct=true&db=eric&AN=EJ928623&site=ehost-live; http://dx.doi.org.ezp2.lib.umn.edu/10.1111/j.1469-7610.2010.02329.x

Sullivan, J. R., & Castro-Villarreal, F. (2013). Special Education Policy, Response to Intervention, and the Socialization of Youth. *Theory Into Practice*, *52*(3), 180-189. doi:10.1080/00405841.2013.804309

Sutton, A. J. (2009). Publication bias. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The Handbook of Research Synthesis and Meta-Analysis* (2nd ed., pp. 435-452). New York: Russel Sage Foundation.

Swanson, H. (1999). Reading research for students with LD: A meta-analysis of intervention outcomes. *Journal Of Learning Disabilities*, *32*(6), 504.

Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition and Instruction*, *8*(4), 351-362. Retrieved from http://www.jstor.org/stable/3233599

Sweller, J., van Merrienboer, Jeroen J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*,*10*(3), 251-296. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=a2h&AN=965008&site=ehost-live

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*(1), 4-28.

U.S. Department of Education, Institute of Education Sciences. National Center for Education Evaluation and Regional Assistance. (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide.* Washington, DC. Retrieved from http://www2.ed.gov/rschstat/research/pubs/rigorousevid/rigorousevid.pdf

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, March). *Beginning Reading intervention report: Fast ForWord®*. Retrieved from http://whatworks.ed.gov

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2010, August). *Adolescent Literacy intervention report: Fast ForWord®*. Retrieved from http://whatworks.ed.gov.

U.S. Department of Education, National Center for Education Statistics. (2014). *The Condition of Education 2014* (NCES 2014-083), Reading Performance. Retrieved from http://nces.ed.gov/fastfacts/display.asp?id=147

U.S. Department of Education, National Center for Education Statistics. (2010). *Teachers' Use of Educational Technology in U.S. Public Schools: 2009* (NCES 2010-040).

U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service, *The U.S.-China E-Language Project: A Study of a Gaming Approach to English Language Learning for Middle School Students.* Washington, D.C., 2011.

van Merriënboer, Jeroen J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*(2), 147-177. doi:10.1007/s10648-005-3951-0

Vaughn, S., Wexler, J., Roberts, G., Barth, A. A., Cirino, P. T., Romain, M. A., . . . Denton, C. A. (2011). Effects of individualized and standardized interventions on middle school students with reading disabilities. Exceptional Children, 77(4), 391-407. Retrieved from http://search.ebscohost.com.ezp1.lib.umn.edu/login.aspx?direct=true&db=aph&AN=63017925&site=ehost-live

Vellutino, F.R., Scanlon, D.M. & Zhang, H. (2007). Identifying reading disability based on response to intervention: Evidence from early intervention research. In S.R. Jimerson, M.K. Burns, & A.M. VanDerHeyden (Eds.) *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 185-211). New York, NY: Springer.

Viechtbauer, W. (2010a). Metafor: Meta-analysis package for R. (Version 1.4-0) [Computer software]. Retrieved from http://CRAN.R-project.org/package=metafor

Viechtbauer, W. (2010b). Conducting meta-analyses in R with the Metafor Package. *Journal of Statistical Software, 36*(03) 1-48.

Walsh, M. E., Madaus, G. F., Raczek, A. E., Dearing, E., Foley, C., An, C., et al. (2014). A new model for student support in high-poverty urban elementary schools: Effects on elementary and middle school academic outcomes. *American Educational Research Journal, 51*(4), 704-737. doi:10.3102/0002831214541669

Wanzek, J., Vaughn, S., Wexler, J., Swanson, E., Edmonds, M. E., & Kim, A.-H. (2006). A synthesis of spelling and reading interventions and their effects on the spelling outcomes of students with LD. *Journal of Learning Disabilities*, 39, 528–543. doi:10.1177/00222194060390060501

Waxman, H., Lin, M., & Michko, G. Learning Point Associates, (2003) *A meta-analysis of the effectiveness of teaching and learning with technology on student outcomes*. Retrieved from http://treeves.coe.uga.edu/edit6900/metaanalysisNCREL.pdf

Wells, J., Lewis, L., National Center for, E. S., & Westat, Inc. Rockville, M.D. (2006). *Internet access in U.S. public schools and classrooms: 1994-2005. highlights. NCES 2007-020.*National Center for Education Statistics. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED494307&site=ehost-live

Wild, M. (2009). Using computer-aided instruction to support the systematic practice of phonological skills in beginning readers. *Journal of Research in Reading, 32*(4), 413-432. doi:10.1111/j.1467-9817.2009.01405.x

Appendix A
Steps in the Search Process

1. Database Abstract Search
    a. Search each set of terms
    b. Combine all sets of terms
    c. Report numbers for each database
    d. Import studies into Refworks
    e. Eliminate duplicates – report number left
2. Abstract Screening
    a. Read titles and abstracts and eliminate articles
    b. Report numbers eliminated based on following hierarchy:
        i. Not an intervention/instruction study (e.g., qualitative, predictive utility of assessments)
        ii. Not reading (e.g., math, science)
        iii. Not CAI
            1. AT
            2. ET
        iv. Not group comparison
            1. Single subject
            2. No comparison group (e.g., pre-post of one group)
            3. Not a primary study (e.g., literature review, meta-analysis)
        v. Not English (i.e., published in a language other than English, CAI in a language other than English)
3. Full-Text Screening
    a. Read full text and eliminate articles
    b. Report numbers eliminated based on following hierarchy:
        i. Exclusion Criteria 1:
            1. Not school-based CAI
            2. Not reading
        ii. Exclusion Criteria 2:
            1. Not primarily delivered through computer
            2. Not interactive
            3. Student-computer ratio greater than 2:1
            4. Not more than 1 session
        iii. Exclusion Criteria 3: Published before 2000
        iv. Exclusion Criteria 4: Participants not in preK-12
        v. Exclusion Criteria 5:

1. Not experimental
2. No comparison group
3. Not able to compute effect size

vi. Exclusion Criteria 6: Article or CAI not in English

vii. Exclusion Criteria 7: Data reported in more than 1 format. More complete data retained.

4. Reference List Review
   a. Review reference lists of included articles for possible studies
   b. Conduct full-text screening of additional references and report number added
5. Search Journals
   a. Search journals represented in included articles using the search terms related to CAI
   b. Conduct full-text screening of additional references and report number added
6. Search IES Reports
   a. Search www.whatworksclearinghouse.com for additional reading CAI intervention reports
   b. Conduct full-text screening of additional references and report number added
7. Search Websites
   a. Conduct a web search to locate websites representing commercially available CAI
   b. Search websites for additional technical reports
   c. Conduct full-text screening of additional references and report number added
8. Contact researchers
   a. Contact researchers of researcher-created or noncommercial CAI to locate additional unpublished studies
   b. Conduct full-text screening of additional references and report number added

Appendix B
Coding Manual

Report Information

Variable	Codes	Description
ESID		Unique number for each study, starting at 1.1. Each additional effect within one report increased by a tenth (e.g., 1.2, 1.3, etc.)
Title		Entire title of report
Author		First two authors' names
PubYear		Year of report publication or year made available if unpublished
PubType		Type of publication
	1	Journal article
	2	Dissertation/Thesis
	3	Book chapter
	4	Technical Report
	5	Conference paper
	6	Other
Journal		Entire title of journal if applicable
Country		Country in which study was conducted
	1	USA
	2	Great Britain/England
	3	Other/not Specified
RefSource		Source of the reference
	1	Database search
	2	Reference list review
	3	Journal search
	4	IES report search
	5	Website search
	6	Personal communication
Funding		Funding for reported study
	1	Government
	2	Private
	3	Other/not specified
NumES		Number of effect sizes included in report
N		Total number of participants included in report
NTxt		Total number of participants in treatment group(s)
NCtl		Total number of participants in control group(s)

Sample Characteristics

Variable	Codes	Description
nTot		Total number of participants within this effect size
Subgroup		Subgroup of a larger sample in report
	0	No
	1	Yes
nTxt		Number of participants in treatment group for this effect size
nCtl		Number of participants in control group for this effect size
SamSel		Type of sample selection
	1	Simple random sampling
	2	Stratified random sampling
	3	Convenience sampling
	4	Entire population
	5	Mix/other
AttTot		Total percentage of attrition reported for this effect size
AttTxt		Percentage of attrition in treatment group for this effect size
AttCtl		Percentage of attrition in control group for this effect size

Participant Demographics

Variable	Codes	Description
SchLvl		Level of school of participants for this effect size
		Grade(s) represented as reported (PreK, K, 1-12)
	1	Grades PreK-Kindergarten
	2	Grades K-6
	3	Grades 6-12
	4	Mixed Levels
%WHT		Percentage of entire sample reported to be White/Caucasian
%BLK		Percentage of entire sample reported to be Black/African American
%HSP		Percentage of entire sample reported to be Hispanic/Latino
%API		Percentage of entire sample reported to be Asian/Pacific Islander
%OTH		Percentage of entire sample reported to be of another or mixed race
Loc		Location of participants/school in study
	1	Urban
	2	Suburban
	3	Rural
	4	Mixed/Other/Not specified
TotSES		Percentage of total sample with low socio-economic status – use entire school if that is all that is provided
TxtSES		Percentage of treatment group with low socio-economic status
TxtCtl		Percentage of control group with low socio-economic status
BoysTot		Total percentage of boys in sample for this effect size
BoysTxt		Percentage of boys in treatment group for this effect size
BoysCtl		Percentage of boys in control group for this effect size

Learner Characteristics

Variable	Codes	Description
ELLTot		Total percentage of sample reported to be English language learners (ELL)
ELLTxt		Percentage of treatment reported to be ELL
ELLCtl		Percentage of control reported to be ELL
LDTot		Total percentage of sample reported to have learning disabilities (LD)
LDTxt		Percentage of treatment reported to have LD
LDCtl		Percentage of control reported to have LD
DCDTot		Total percentage of sample reported to have developmental cognitive disabilities (DCD)
DCDTxt		Percentage of treatment reported to have DCD
DCDCtl		Percentage of control reported to have DCD
ASDTot		Total percentage of sample reported to have autism spectrum disorders (ASD)
ASDTxt		Percentage of treatment reported to have ASD
ASDCtl		Percentage of control reported to have ASD
EBDTot		Total percentage of sample reported to have emotional/behavioral disorders (EBD)
EBDTxt		Percentage of treatment reported to have EBD
EBDCtl		Percentage of control reported to have EBD
DHHTot		Total percentage of sample reported to be deaf or hard of hearing (DHH)
DHHTxt		Percentage of treatment reported to be DHH
DHHCtl		Percentage of control reported to be DHH
VITot		Total percentage of sample reported to be visually impaired
VITxt		Percentage of treatment reported to be visually impaired
VICtl		Percentage of control reported to be visually impaired
AtRTot		Total percentage of sample reported to be at-risk academically
AtRTxt		Percentage of treatment reported to be at-risk academically
AtRCtl		Percentage of control reported to be at-risk academically
OthTot		Total percentage of sample reported to have other/undisclosed disabilities
OthTxt		Percentage of treatment reported to have other/undisclosed disabilities
OthCtl		Percentage of control reported to have other/undisclosed disabilities

Intervention Information

Variable	Codes	Description
IntLoc		Location of the intervention
	1	In core literacy classroom
	2	Separate room
	3	Home
	4	Mixed
	5	Other/not specified
LitCom		Component of literacy
	1	Supplemental to core literacy (additional time)
	2	Integrated in core literacy (no additional time)
	3	Mixed/other/not specified
Control		Type of control/comparison group
	1	No treatment
	2	Traditional intervention
	3	Non-reading CAI
	4	Mixed
	5	Other/not specified
SessLgth		Length of each session (min)
TotLgth		Total length of intervention (hours)
IntLgthD		Total number of days of the intervention
IntLgthW		Total number of weeks of the intervention
SessWks		Sessions per week
Teacher		Teacher who implemented the intervention
	1	Licensed teacher
	2	Researchers
	3	Other school staff
	4	Not specified

CAI Characteristics

Variable	Codes	Description
CAI Name		Entire name of CAI used for this effect size
Soft		Commercially available software
	0	No
	1	Yes
PhA		Program addresses phonemic awareness
	0	No
	1	Yes
Phn		Program addresses phonics
	0	No
	1	Yes
Flu		Program addresses fluency
	0	No
	1	Yes

CAI Characteristics (cont.)

Variable	Codes	Description
Voc		Program addresses vocabulary
	0	No
	1	Yes
Com		Program addresses comprehension
	0	No
	1	Yes
GA		Program uses guided activity
	0	No
	1	Yes
CF		Program uses corrective feedback
	0	No
	1	Yes
EF		Program uses elaborate feedback
	0	No
	1	Yes
PA		Program allows users to go at their own pace
	0	No
	1	Yes
PT		Program uses pre-training to activate background knowledge
	0	No
	1	Yes
RK		Program keeps records of student data
	0	No
	1	Yes
CAISt		Structure of CAI
	1	Linear
	2	Adaptive
	3	Not reported

Outcome Characteristics

Variable	Codes	Description
DVname		Name of measure for this effect size
Stand		Standardized assessment procedures and scoring
	0	No
	1	Yes
Broad		Broad measure of reading (two or more areas addressed, including comprehension)
	0	No
	1	Yes
SubTst		Subtest of a larger test
	0	No
	1	Yes

Outcome Characteristics (cont.)

Variable	Codes	Description
RC		Researcher created measure
	0	No
	1	Yes
MultDV		One of multiple measures in this report
	0	No
	1	Yes
APhA		Phonemic awareness assessed
	0	No
	1	Yes
Aphn		Phonics assessed
	0	No
	1	Yes
Aflu		Fluency assessed
	0	No
	1	Yes
Avoc		Vocabulary assessed
	0	No
	1	Yes
Acomp		Comprehension assessed
	0	No
	1	Yes
Align		Measure aligned with at least one CAI component
	0	No
	1	Yes

Results

Variable	Codes	Description
PreTxtM		Pretest mean of treatment group
PreTxtSD		Pretest standard deviation of treatment group
PreCtlM		Pretest mean of control group
PreCtlSD		Pretest standard deviation of control group
TxtM		Posttest mean of treatment group
TxtSD		Posttest standard deviation of treatment group
CtlM		Posttest mean of control group
CtlSD		Posttest standard deviation of control group
ES		Computed effect size (Hedges g)
ConfLow		Lower limit of confidence interval
ConfUp		Upper limit of confidence interval
OtherES		Other values used to compute effect size (e.g., t-test)
IndES		Independent effect size
	0	No
	1	Yes
StudyES		Reported effect size

Results (cont.)

Variable	Codes	Description
AdjM		Adjusted mean used to compute effect size
	0	No
	1	Yes
Gain		Gain score used to compute effect size
	0	No
	1	Yes
Mot		Motivation if assessed
	0	Negative
	1	Positive
ArchData		Study used archived data to determine results
	0	No
	1	Yes

Quality Indicators (questions verbatim from Gersten et al., 2005)

Variable	Codes	Description
Abil		Was sufficient information provide to determine/confirm whether the participants demonstrated the disability(ies) or difficulties presented?
	0	No
	1	Yes
CompG		Were appropriate procedures used to increase the likelihood that relevant characteristics of participants in the sample were comparable across conditions?
	0	No
	1	Yes
CompInt		Was sufficient information given characterizing the interventionists or teachers provided? Did it indicate whether they were comparable across conditions?
	0	No
	1	Yes
Att		Was data available on attrition rates among intervention samples? Was severe overall attrition documented? If so, is attrition comparable across samples? Is overall attrition less than 30%
	0	No
	1	Yes
IntDes		Was the intervention clearly described and specified?
	0	No
	1	Yes
Fid		Was the fidelity of implementation described and assessed?
	0	No
	1	Yes

Quality Indicators (cont.)

CntDes		Was the nature of services provided in comparison conditions described?
	0	No
	1	Yes
MultMeas		Were multiple measures used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalized performance?
	0	No
	1	Yes
MeasTm		Were outcomes for capturing the intervention's effect measured at the appropriate times?
	0	No
	1	Yes
Data		Were the data analysis techniques appropriately linked to key research questions and hypotheses? Were they appropriately linked to the unit of analysis in the study?
	0	No
	1	Yes
ESR		Did the research report include not only inferential statistics but also effect size calculations?
	0	No
	1	Yes
RCT		Randomized control trial
	0	No
	1	Yes
Score		Total quality score (add above)
	0	No
	1	Yes
Score%		Total quality score divided by 12
	0	No
	1	Yes

Appendix C
Descriptive Summary of Included Reports

Table C1

Results of CAI used in Elementary (PreK-6th Grade)

CAI Name	Pub Type[a]	N	Trt *n*	Ctl *n*	Ctl Type	Grade	Div[b]	Low SES[c]	ELL[d]	IEP[e]	At-Risk	# of ES[f]	Mean ES (*g*)	95% CI	QI
Academy of Reading															
Campuzano et al. (2009)	TR	899	495	404	No Trt	4	83%	65%	x	x	53%	1	-0.16*	[-0.29, -0.02]	58%
Greenlee (2001)	DI	128	64	64	x	2-6	73%	x	x	x	100%	6	-0.37*	[-0.71, -0.02]	33%
Breakthrough to Literacy															
Woodward (2005)	DI	149	85	64	No Trt	K	x	20%	x	x	x	1	+0.25	[-0.07, +0.58]	25%
DaisyQuest and Daisy's Castle															
Lonigan et al. (2003)	JN	41	20	21	No Trt	PreK	90%	100%	x	x	x	8	+0.41	[-0.20, 1.02]	75%
Mathes et al. (2001)	JN	85	43	42	No Trt	1	54%	x	x	12% LD	100%	9	-0.10	[-0.52, +0.32]	75%
Mitchell & Fox (2001)	JN	48	24	24	Trad Int	K-1	9%	x	x	x	100%	10	-0.46	[-1.02, +0.11]	75%
Mitchell & Fox (2001)	JN	48	24	24	No Trt	K-1	9%	x	x	x	100%	10	+0.69*	[+0.11, +1.26]	75%
Destination Reading															
Campuzano et al. (2009)	TR	742	448	294	No Trt	1	65%	71%	x	x	35%	5	-0.33	[-0.05, +0.24]	58%
Earobics															
Gale (2006)	DI	26	25	25	No Trt	K-1	51%	73%	x	x	100%	7	+1.41*	[+0.81, +2.02]	67%
Gillam et al. (2008)	JN	105	53	52	Trad Int	K-4	54%	x	x	100% LI	x	2	-0.17	[-0.55, +0.21]	92%
Gillam et al. (2008)	JN	106	53	53	No Trt	K-4	54%	x	x	100% LI	x	2	+0.29	[-0.09, +0.67]	92%

Table C1 (cont.)

Results of CAI used in Elementary (PreK-6th Grade)

Earobics (cont.)															
Pokorni et al. (2004)	JN	34	16	18	Trad Int	PreK-5	79%	50%	x	100% LI	100%	5	-0.00	[-0.66, +0.70]	67%
Rehmann (2005)	DI	66	30	36	No Trt	K-1	49%	58%	x	x	x	7	+0.01	[-0.47, +0.49]	42%
Valliath (2002)	DI	30	15	15	Non Read CAI	1	x	x	x	x	x	6	+0.44	[-0.27, +1.15]	50%
English Language Learners Instructional System (ELLIS)															
Beaird (2007)	DI	61	42	19	No Trt	3-5	100%	66%	100%	66% LD 33% DCD 16% VI	x	2	-0.76	[-1.65, +0.13]	50%
Fast ForWord															
Bjorn & Leppanen (2013)	JN	24	13	11	No Trt	5	x	x	100%	x	100%	4	-0.18	[-0.96, +0.60]	42%
Cohen et al. (2005)	JN	50	23	27	No Trt	K-5	x	x	x	100% LI	x	4	+0.29	[-0.26, +0.84]	33%
Gillam et al. (2008)	JN	103	51	52	Trad Int	K-4	54%	x	x	100% LI	x	2	-0.06	[-0.44, +0.32]	92%
Gillam et al. (2008)	JN	104	51	53	Non Read CAI	K-4	54%	x	x	100% LI	x	2	+0.41*	[+0.03, +0.80]	92%
Hook et al. (2001)	JN	20	11	9	Trad Int	K-6	x	x	x	x	100%	8	-0.31	[-1.16, +0.55]	25%
Hook et al. (2001)	JN	20	11	11	No Trt	K-6	x	x	x	x	100%	8	+0.12	[-0.69, +0.93]	25%
Pokorni et al. (2004)	JN	38	20	18	Trad Int	1-3	79%	52%	x	100% LI	100%	5	-0.45	[-1.09, +0.18]	67%
Rouse & Krueger (2004)	JN	463	244	219	No Trt	3-6	x	70%	x	x	100%	2	+0.90*	[+0.71, +1.10]	42%
Troia (2004)	JN	168	90	78	No Trt	1-6	x	x	57%	x	x	6	-0.12	[-0.56, +0.31]	67%

Table C1 (cont.)

Results of CAI used in Elementary (PreK-6th Grade)

CAI Name	Pub Type[a]	N	Trt *n*	Ctl *n*	Ctl Type	Grade	Div[b]	Low SES[c]	ELL[d]	IEP[e]	At-Risk	# of ES[f]	Mean ES (*g*)	95% CI	QI
Fast ForWord (cont.)															
Troia & Whitney (2003)	JN	37	25	12	No Trt	1-6	33%	59%	x	x	100%	6	+0.44	[-0.27, +1.15]	42%
First 4000 Words															
Fehr (2011)	DI	43	22	21	No Trt	2-4	x	70%	x	x	100%	4	+0.71*	[+0.11, +1.32]	50%
Fehr (2011)	DI	192	96	96	No Trt	2-4	x	49%	38%	x	100%	4	+0.02	[-0.42, +.46]	42%
Headsprout Early Reading															
Campuzano et al. (2009)	TR	1,079	574	505	No Trt	1	19%	34%	x	x	23%	1	-0.09	[-0.21, +0.03]	58%
Huffstetter et al. (2010)	JN	62	31	31	Non Read CAI	PreK	100%	x	51%	x	x	1	+1.37	[+0.82, +1.92]	83%
Kreskey (2012)	DI	102	51	51	No Trt	K	48%	50%	6%	x	100%	2	-0.46*	[-0.85, -0.07]	58%
Kreskey (2012)	DI	152	76	76	No Trt	1	37%	45%	1%	x	100%	2	+0.07	[-0.25, +0.39]	58%
Imagine Learning English															
Longberg (2012)	DI	273	134	139	No Trt	K	x	64%	24%	x	x	2	+0.12	[-0.11, +0.36]	83%
Intelligent Tutoring System for the Structure Strategy (ITSS)															
Wijekumar et al. (2012)	JN	130	64	66	No Trt	4	x	x	x	x	x	1	+0.32	[-0.02, +0.66]	58%
Wijekumar et al. (2013)	JN	850	449	401	x	4-5	x	x	x	x	x	2	+0.34*	[+0.21, +0.48]	58%
Leap Track															
Campuzano et al. (2009)	TR	1,274	665	609	No Trt	1	67%	61%	x	x	50%	1	-0.03	[-0.14, +0.08]	58%

Table C1 (cont.)

Results of CAI used in Elementary (PreK-6th Grade)

CAI Name	Pub Type[a]	N	Trt *n*	Ctl *n*	Ctl Type	Grade	Div[b]	Low SES[c]	ELL[d]	IEP[e]	At-Risk	# of ES[f]	Mean ES (*g*)	95% CI	QI
Lexia															
Gale (2006)	DI	24	12	12	No Trt	K-1	54%	73%	x	x	100%	7	+0.66	[-0.13, +1.45]	67%
Macaruso et al. (2006)	JN	167	83	84	No Trt	1	x	52%	x	x	18%	4	+0.06	[-0.24, +0.37]	67%
Macaruso & Rodman (2011a)	JN	66	29	37	Non Read CAI	K	x	x	100%	x	x	14	+0.60*	[+0.11, 1.10]	83%
Macaruso & Rodman (2011b)	JN	38	19	19	No Trt	PreK	65%	42%	x	x	x	4	+0.57	[-0.07, +1.21]	50%
Macaruso & Rodman (2011b)	JN	66	37	19	No Trt	K	56%	74%	x	x	100%	6	+0.48	[-0.05, +1.02]	50%
Macaruso & Walker (2008)	JN	71	26	45	No Trt	K	x	50%	x	x	33%	13	+0.20	[-0.28, 0.68]	58%
Literacy CD-ROMs															
Cohen et al. (2005)	JN	54	27	27	No Trt	K-5	x	x	x	100% LI	x	4	+0.14	[-0.38, +0.67]	33%
Multimedia Talking Books															
Chera & Wood (2003)	JN	30	15	15	No Trt	PreK-K	x	x	x	x	x	7	+0.27	[-0.43, +0.98]	25%
Project LISTEN Reading Tutor															
Mostow et al. (2003)	JN	178	88	90	No Trt	1-4	x	10%	x	x	x	7	+0.37*	[+0.07, +0.67]	67%
Read Naturally															
Arvans (2010)	DI	82	39	43	No Trt	2-4	73%	62%	x	x	100%	7	+0.05	[-0.38, +0.48]	83%
Christ & Davie (2009)	TR	106	53	53	No Trt	3	57%	60%	23%	x	100%	7	+0.09	[-0.29, +0.47]	92%

Table C1 (cont.)

Results of CAI used in Elementary (PreK-6th Grade)

CAI Name	Pub Type[a]	N	Trt *n*	Ctl *n*	Ctl Type	Grade	Div[b]	Low SES[c]	ELL[d]	IEP[e]	At-Risk	# of ES[f]	Mean ES (*g*)	95% CI	QI
Reader's Interactive Teaching Assistant (RITA)															
Nicolson et al. (2000)	JN	52	16	36	Trad Int	K-1	x	x	x	x	100%	2	+0.51	[-0.08, +1.10]	42%
Nicolson et al. (2000)	JN	61	16	45	No Trt	K-1	x	x	x	x	100%	2	+0.75	[+0.18, +1.33]	42%
Reading Plus															
Rasinski et al. (2011)	JN	2,536	1,558	978	No Trt	4-5	90%	x	3%	6% LD	x	48	+0.11*	[+0.01, +0.20]	33%
Reutzel et al. (2012)	JN	80	40	40	Trad Int	3	97%	x	14%	15% IEP	100%	2	+0.72*	[+0.27, +1.17]	50%
Shelley-Tremblay & Eyer (2009)	JN	77	45	32	No Trt	2	9%	x	x	x	x	3	+0.33	[-0.13, +0.78]	50%
Reading Upgrade															
Cole & Hilliard (2006)	JN	36	18	18	Trad Int	3	36%	97%	x	x	100%	2	+2.29*	[+1.46, +3.12]	67%
Rhyme & Analogy CD-ROM															
Wild (2009)	JN	87	44	43	Trad Int	K	x	x	x	x	x	8	+0.32	[-0.10, +0.74]	67%
Wild (2009)	JN	84	44	40	No Tret	K	x	x	x	x	x	8	+0.30	[-0.13, +0.72]	
Speech Recognition-Based Program (SRBP)															
Higgins & Raskind (2004)	JN	44	28	16	No Trt	1-12	20%	x	x	100% LD	100%	6	-0.26	[-0.88, +0.35]	42%
Success Maker															
Gatti & Miller (2011)	TR	619	352	267	No Trt	3	38%	36%	x	x	24%	6	+0.16*	[+0.00, +0.32]	75%
Gatti & Miller (2011)	TR	641	342	299	No Trt	5	51%	45%	x	x	20%	5	+0.07	[-0.09, +0.22]	75%

Table C1 (cont.)

Results of CAI used in Elementary (PreK-6th Grade)

CAI Name	Pub Type[a]	N	Trt *n*	Ctl *n*	Ctl Type	Grade	Div[b]	Low SES[c]	ELL[d]	IEP[e]	At-Risk	# of ES[f]	Mean ES (*g*)	95% CI	QI
Waterford Early Reading Program (WERP)															
Campuzano et al. (2009)	TR	1,115	689	466	No Trt	1	93%	47%	x	x	31%	1	+0.03	[-0.08, +0.15]	58%
Cassady & Smith (2004)	JN	88	26	62	No Trt	K	x	x	x	x	x	2	+0.50*	[+0.04, +0.96]	25%
Cassady & Smith (2005)	JN	93	46	47	No Trt	1	x	11%	x	x	25%	4	+0.56*	[+0.15, +0.97]	58%
Fischel et al. (2007)	JN	357	172	185	Trad Int	PreK	83%	100%	x	x	x	7	-0.17	[-0.38, +0.03]	75%
Fischel et al. (2007)	JN	322	172	150	No Trt	PreK	83%	100%	x	x	x	7	+0.12	[-0.10, +0.34]	75%
Hansen et al. (2004)	TR	848	442	406	No Trt	K	x	x	x	x	95%	3	+1.25*	[+1.09, +1.41]	75%
Hansen et al. (2004)	TR	1,015	577	438	No Trt	1	x	x	x	x	95%	7	-2.16*	[-2.34, -1.98]	75%
Hecht & Close (2002)	JN	76	42	34	No Trt	K	x	x	x	x	x	7	+0.56*	[+0.10, +1.03]	33%
Johnson et al. (2010)	JN	183	58	54	No Trt	PreK	x	x	x	x	x	3	+0.44*	[+0.03, +0.84]	58%
Paterson et al. (2003)	JN	108	59	49	No Trt	K-1	x	67%	x	x	x	5	-0.30	[-0.68, +0.08]	33%
Powers & Price-Johnson (2006)	TR	1,545	344	1,211	x	K	81%	x	x	x	23%	8	+0.30*	[+0.18, +0.43]	33%
Stevenson (2006)	TR	198	86	112	No Trt	PreK	x	x	57%	x	x	4	+0.43*	[+0.13, +0.73]	50%
Tracey & Young (2007)	JN	265	151	114	No Trt	K	x	77%	x	x	100%	3	+0.28*	[+0.03, +0.52]	50%

Note. Ctl = Control *n* ; Trt = Treatment *n;* Ctl Type = control type: no treatment or traditional intervention; x = not reported; CI = confidence interval, QI = quality indicators met (percentage)
[a]PubType = Publication Type, JN = journal article, TR = technical report, DI = dissertation; [b]Div = Diversity, or the reported non-white percentage of sample; [c]Low SES is the percentage of the sample reported to be receiving free-or-reduced lunch; [d]ELL = English Language Learner percentage of sample; [e]IEP = Individualized Education Plan, or the percentage of students receiving special education services, LD = Learning Disabilities, DCD = Developmental Cognitive Disabilities, LI = Language Impaired, VI = Visually Impaired; [f]Number of effect sizes for that program in each report
*Statistically significant (95% CI does not include zero)

Table C2

Results of CAI used in Middle and High School (6th-12th Grade)

CAI Name	Pub Type[a]	N	Trt *n*	Ctl *n*	Ctl Type	Grade	Div[b]	Low SES[c]	ELL[d]	IEP[e]	At-Risk	# of ES[f]	Mean ES (*g*)	95% CI	QI
Academy of Reading															
Greenlee (2001)	DI	16	8	8	x	6	73%	x	x	x	100%	1	-0.53	[-1.47, +0.42]	33%
Compass Learning Odyssey															
Gillard (2010)	DI	1,223	612	611	No Trt	6	51%	47%	x	x	x	1	+0.02	[-0.09, +0.13]	17%
Computer-Assisted Collaborative Strategic Reading (CACSR)															
Kim et al. (2006)	JN	34	16	18	No Trt	6-8	56%	50%	x	82% LD 18% IEP	x	5	+0.83*	[+0.14, +1.52]	58%
Fast ForWord															
Given et al. (2008)	JN	25	12	13	No Trt	6-8	49%	x	x	x	100%	5	+0.01	[-0.75, +0.77]	67%
Fast ForWord + Success Maker															
Given et al. (2008)	JN	28	15	13	No Trt	6-8	49%	x	x	x	100%	10	+0.26	[-0.50, +1.02]	67%
Independent Silent Reading Modules (ISR)															
Cuevas et al. (2012)	JN	62	24	38	Trad Int	10	x	x	x	x	x	2	+0.04	[-0.47, +0.54]	33%
Cuevas et al. (2012)	JN	67	24	43	No Trt	10	x	x	x	x	x	2	+0.58*	[+0.08, +1.09]	33%

Table C2 (cont.)

Results of CAI used in Middle and High School (6th-12th Grade)

CAI Name	Pub Type[a]	N	Trt *n*	Ctl *n*	Ctl Type	Grade	Div[b]	Low SES[c]	ELL[d]	IEP[e]	At-Risk	# of ES[f]	Mean ES (*g*)	95% CI	QI
Information and Communication Technology (ICT)															
Brooks et al. (2006)	JN	138	67	71	No Trt	6	x	x	x	x	x	1	-0.22	[-0.55, +0.11]	33%
Khan & Gorard (2012)	JN	665	319	346	No Trt	6	x	21%	x	x	x	1	-0.28*	[-0.43, -0.12]	58%
Lexia Strategies for Older Students (S.O.S.)															
Macaruso & Rodman (2009)	JN	42	27	15	No Trt	6-7	17%	x	x	x	100%	6	-0.26	[-0.89, +0.36]	75%
Rupprecht (2003)	DI	20	10	10	Trad Int	6-8	40%	x	30%	x	100%	4	-0.02	[-0.88, +0.83]	33%
Merit Software															
Jones et al. (2004)	JN	151	116	35	No Trt	6, 8	x	x	x	x	x	2	+0.99*	[+0.23, +1.75]	42%
Peabody Literacy Lab (Read 180 Software)															
Hasselbring & Goin (2004)	JN	122	60	62	No Trt	6-8	x	x	x	x	x	6	-0.08	[-0.44, +0.27]	17%
Reading Plus															
Rasinski et al. (2011)	JN	6,725	2,500	4,225	No Trt	6-8	90%	x	3%	6% LD	x	72	+0.22*	[+0.17, +0.28]	33%
Rasinski et al. (2011)	JN	6,882	1,700	5,182	No Trt	9-10	90%	x	3%	6% LD	x	40	+0.08*	[+0.01, +0.14]	33%

Table C2 (cont.)

Results of CAI used in Middle and High School (6^{th}-12^{th} Grade)

CAI Name	Pub Type[a]	N	Trt *n*	Ctl *n*	Ctl Type	Grade	Div[b]	Low SES[c]	ELL[d]	IEP[e]	At-Risk	# of ES[f]	Mean ES (*g*)	95% CI	QI
Success Maker															
Gatti & Miller (2011)	TR	453	254	199	No Trt	7	47%	53%	x	x	32%	5	+0.26*	[+0.08, +0.45]	75%
Given et al. (2008)	JN	24	11	13	No Trt	6-8	49%	x	x	x	100%	5	+0.09	[-0.65, +0.82]	67%
The Forgotten World															
Green et al. (2011)	TR	3,592	1,180	1,782	No Trt	8	100%	x	100%	x	x	3	+0.03	[-0.04, +0.09]	33%

Note. Ctl = Control *n* ; Trt = Treatment *n;* Ctl Type = control type: no treatment or traditional intervention; x = not reported; CI = confidence interval, QI = quality indicators met (percentage)
[a]PubType = Publication Type, JN = journal article, TR = technical report, DI = dissertation; [b]Div = Diversity, or the reported non-white percentage of sample; [c]Low SES is the percentage of the sample reported to be receiving free-or-reduced lunch; [d]ELL = English Language Learner percentage of sample; [e]IEP = Individualized Education Plan, or the percentage of students receiving special education services, LD = Learning Disabilities, DCD = Developmental Cognitive Disabilities, LI = Language Impaired, VI = Visually Impaired; [f]Number of effect sizes for that program in each report
*Statistically significant (95% CI does not include zero

Appendix D
Summary of CAI Characteristics

Table D1

Summary of CAI Characteristics

	Intervention			Areas of Reading Addressed by CAI					Instructional Components Reported						Technology Structure				Dependent Variables Addressed							
CAI Name	Grade	Inst Lev	Tot Dur	P A	P H	F L	V O	C M	G A	C F	E F	P C	P T	R K	L N	A D	W B	C A	B R	P A	P H	F L	V O	C M	ES Range (*g*)	QI[1]
Academy of Reading	4-6		10	+	+	+		+		+		+		+	+		+	+	+				+	+	-0.53- -0.16 (*k*=7)	33-58%
Breakthrough to Literacy	K	Int	40	+											+					+					+0.25 (*k*=1)	25%
Compass Learning Odyssey	6				+		+			+		+	+	+	+		+	+	+						+0.02 (*k*=1)	17%
Computer-Assisted Collaborative Strategic Reading (CACSR)	6-8	Int	18					+	+	+		+		+	+									+	+0.83 (*k*=5)	58%
DaisyQuest and Daisy's Castle	PreK-1	Sup	5-9	+						+		+			+			+		+	+			+	-0.46- +0.69 (*k*=27)	75%
Destination Reading	1		11		+			+						+	+			+	+						-0.33 (*k*=5)	58%
Earobics	PreK-5	Sup and Int	8-50	+	+					+	+	+		+	+		+	+		+	+	+		+	-0.17- +1.41 (*k*=37)	42-92%

Table D1 (cont.)

Summary of CAI Characteristics

	Intervention			Areas of Reading Addressed by CAI					Instructional Components Reported						Technology Structure				Dependent Variables Addressed							
CAI Name	Grade	Inst Lev	Tot Dur	PA	PH	FL	VO	CM	GA	CF	EF	PC	PT	RK	LN	AD	WB	CA	BR	PA	PH	FL	VO	CM	ES Range (*g*)	QI[1]
English Language Learners Instructional System (ELLIS)	3-5	Sup	26		+		+						+		+		+	+			+			+	-0.76 (*k*=2)	50%
Fast ForWord	K-8	Sup and Int	33-88	+						+				+	+		+	+		+	+			+	-0.45-+0.90 (*k*=52)	25-92%
Fast ForWord + Success Maker	6-8	Int	88	+	+		+	+						+	+		+	+		+	+			+	+0.26 (*k*=10)	67%
First 4000 Words	2-4	Sup	2-9				+		+	+		+		+	+		+	+					+		+0.02-+0.71 (*k*=8)	42-50%
The Forgotten World	8									+	+	+		+	+		+	+	+						+0.03 (*k*=3)	33%
Headsprout Early Reading	PreK-1	Sup	7-20	+	+	+	+	+		+		+		+	+		+	+	+	+	+	+	+	+	-0.46-+1.37 (*k*=6)	58-53%
Imagine Learning English	K	Int	17		+	+	+	+						+	+		+	+			+	+	+		+0.12 (*k*=2)	83%
Independent Silent Reading Modules (ISR)	10	Int	14					+				+			+				+						+0.04-+0.58 (*k*=2)	33%
Information and Communication Technology	6		10		+					+		+			+				+						-0.22 (*k*=1)	33%

Table D1 (cont.)

Summary of CAI Characteristics

	Intervention			Areas of Reading Addressed by CAI					Instructional Components Reported						Technology Structure				Dependent Variables Addressed							
CAI Name	Grade	Inst Lev	Tot Dur	PA	PH	FL	VO	CM	GA	CF	EF	PC	PT	RK	LN	AD	WB	CA	BR	PA	PH	FL	VO	CM	ES Range (*g*)	QI[1]
DaisyQuest and Daisy's Castle	6		18		+	+	+	+		+		+		+						+	+				-0.28 (*k*=1)	58%
Intelligent Tutoring System for the Structure Strategy (ITSS)	4-5	Int	15					+	+	+	+	+		+		+	+	+						+	+0.32- +0.34 (*k*=3)	58%
Leap Track	1		9	+	+	+	+	+				+		+					+						-0.03 (*k*=1)	58%
Lexia	PreK-8	Int and Sup	8-32	+	+	+	+	+	+	+		+		+		+	+	+	+	+	+	+	+	+	+0.06- +0.66 (*k*=58)	33-83%
Literacy CD-ROMs	K-5	Sup	27	+	+															+	+				+0.14 (*k*=4)	33%
Merit Software	6, 8	Int	6				+	+		+		+		+	+		+	+					+	+	+0.99 (*k*=2)	42%
Multimedia Talking Books	PreK-K		17	+								+			+					+	+				+0.27 (*k*=7)	25%
Peabody Literacy Lab (Read 180 Software)	6-8	Sup	85		+			+	+	+	+	+	+	+	+		+	+			+		+	+	-0.08 (*k*=6)	17%
Project LISTEN Reading Tutor	1-4	Sup	19			+			+	+		+		+						+	+	+	+		+0.37 (*k*=7)	67%

Table D1 (cont.)

Summary of CAI Characteristics

	Intervention			Areas of Reading Addressed by CAI					Instructional Components Reported						Technology Structure				Dependent Variables Addressed							
CAI Name	Grade	Inst Lev	Tot Dur	PA	PH	FL	VO	CM	GA	CF	EF	PC	PT	RK	LN	AD	WB	CA	BR	PA	PH	FL	VO	CM	ES Range (*g*)	QI[1]
Read Naturally	2-4	Sup	9-18			+	+	+	+	+		+		+	+		+	+			+	+	+	+	+0.05-+0.09 (*k*=14)	83-92%
Reader's Interactive Teaching Assistant (RITA)	K-1	Int	10		+	+		+							+				+						+0.51-+0.75 (*k*=2)	42%
Reading Plus	2-10	Int and Sup	15-53			+	+	+		+		+		+		+	+	+	+		+		+		+0.11-+0.72 (*k*=53)	33-50%
Reading Upgrade	3	Sup	3	+	+			+	+				+		+		+	+	+		+				+2.29 (*k*=2)	67%
Rhyme & Analogy CD-ROM	K	Sup	2	+	+					+		+			+			+		+		+			+0.30-+0.32 (*k*=8)	67%
Speech Recognition-Based Program	1-12		14	+	+					+		+								+	+				-0.26 (*k*=6)	42%
Success Maker	3-8	Int	19-88			+	+	+						+			+			+	+	+	+	+	+0.07-+0.16 (*k*=16)	67-75%
Waterford Early Reading Program	PreK-1	Int and Sup	9-60	+	+	+	+	+		+		+		+	+		+	+	+	+	+	+	+	+	-2.16-+1.25 (*k*=61)	25-75%

Note: Grd = Grade; Inst Level = instructional level; Sup = supplemental to core instruction; Int = integrated in core instruction; Tot Dur = total duration (hours); PA = phonemic/phonological awareness; PH = phonics; FL = fluency; VO = vocabulary; CM = comprehension; GA = guided activity; CF = corrective feedback; EF = elaborate feedback; PC = pacing; PT = pretraining; RK = record keeping; LN = linear; AD = adaptive; WB = web-based; CA = commercially available; BR = broad reading; ES # = total number of reported effects; QI = quality indicator percentage range

Appendix E
Results of Categorical Moderator Analyses

Table E1

Dataset A: Categorical Moderator Analysis of Study Characteristics (k = 263)

Descriptor	*k*	*g*	SE	Var	95% CI	Q_w	I^2	Q_B
Publication Type							75%	8.03
Dissertation	50	+0.39	0.14	0.020	[0.02, 0.11]	168.00*	71%	
Journal	176	+0.25	0.07	0.006	[0.10, 0.39]	556.59*	69%	
Technical Report	37	-0.18	0.16	0.025	[-0.49, 0.13]	8246.81*	99%	
Publication Year							53%	4.25
2000-2004	83	+0.03	0.11	0.012	[-0.18, 0.24]	8211.07*	99%	
2005-2009	113	+0.31	0.09	0.009	[0.13, 0.49]	355.05*	68%	
2010-2013	67	+0.26	0.12	0.014	[0.03, 0.50]	234.89*	72%	
Funding Source							0%	0.90
Government	63	+0.11	0.12	0.015	[-0.14, 0.35]	148.83*	58%	
Private	52	+0.25	0.14	0.019	[-0.02, 0.52]	211.54*	76%	
Other/Not Specified	148	+0.24	0.08	0.007	[0.08, 0.40]	8801.16*	98%	
Country of Study							0%	0.56
USA	234	+0.22	0.07	0.004	[0.09, 0.35]	9221.46*	97%	
Great Britain	22	+0.23	0.22	0.047	[-0.19, 0.66]	45.74*	54%	
Other	7	-0.07	0.39	0.149	[-0.83. 0.68]	6.28	4%	
Location of Study							10%	3.33
Urban	76	+0.04	0.11	0.013	[-0.19, 0.26]	8485.41*	99%	
Suburban	36	+0.27	0.17	0.028	[-0.06, 0.60]	109.85*	68%	
Rural	32	+0.24	0.18	0.032	[-0.11, 0.59]	79.55*	61%	
Mixed/Not Specified	119	+0.30	0.09	0.008	[0.12, 0.47]	387.15*	70%	
Experimental Design							76%	4.19
RCT	176	+0.30	0.08	0.006	[0.15, 0.45]	620.11*	72%	
Quasi-Experimental	87	+0.04	0.11	0.011	[-0.17, 0.24]	8430.97*	99%	

Note: All effects used in this analysis
*$p<0.005$

Table E2

Dataset B: Categorical Moderator Analysis of Study Characteristics (k = 217)

Descriptor	*k*	*g*	SE	Var	95% CI	Q_w	I^2	Q_B
Publication Type							60%	4.93
Dissertation	8	-0.09	0.11	0.012	[-0.30, 0.13]	26.06*	73%	
Journal	193	+0.14	0.02	0.000	[0.11, 0.18]	799.23*	76%	
Technical Report	16	+0.09	0.05	0.002	[-0.01, 0.19]	30.80	51%	
Publication Year							86%	14.34*
2000-2004	15	-0.15	0.08	0.006	[-0.31, 0.00]	27.71	41%	
2005-2009	24	+0.22	0.07	0.005	[0.08, 0.37]	21.05	0%	
2010-2013	178	+0.14	0.02	0.000	[0.10, 0.17]	799.74*	78%	
Funding Source							0%	1.66
Government	30	+0.05	0.07	0.004	[-0.07, 0.18]	31.67	8%	
Private	19	+0.11	0.05	0.002	[0.02, 0.21]	41.31*	56%	
Other/Not Specified	168	+0.14	0.02	0.000	[0.10, 0.17]	787.97*	79%	
Country of Study								
USA	217	+0.13	0.02	0.000	[0.10, 0.16]	870.63*	75%	
Location of Study							61%	7.62
Urban	3	+0.28	0.13	0.016	[0.03, 0.52]	4.10	51%	
Suburban	8	-0.09	0.11	0.012	[-0.30, 0.13]	26.06*	73%	
Rural	1	+0.56	0.28	0.077	[0.02, 1.10]	0	0%	
Mixed/Not Specified	205	+0.13	0.02	0.000	[0.10, 0.16]	828.11*	75%	
Experimental Design							38%	1.63
RCT	50	+0.09	0.04	0.001	[0.02, 0.16]	72.29	32%	
Quasi-Experimental	167	+0.14	0.02	0.000	[0.10, 0.18]	788.51*	79%	

Note: All effects included in analysis.
*p<0.005

Table E3

Dataset C: Categorical Moderator Analysis of Study Characteristics (k = 52)

Descriptor	*k*	*g*	SE	Var	95% CI	Q_w	I^2	Q_B
Publication Type							0%	0.01
Dissertation	5	-0.03	0.24	0.058	[-0.50, 0.45]	7.13	44%	
Journal	47	-0.05	0.06	0.004	[-0.16, 0.07]	152.01	70%	
Technical Report	-	-	-	-	-	-	-	
Publication Year							79%	9.52
2000-2004	24	-0.20	0.09	0.009	[-0.38, -0.02]	29.08	21%	
2005-2009	25	-0.01	0.07	0.005	[-0.16, 0.13]	98.13*	76%	
2010-2013	3	+0.50	0.22	0.047	[0.08, 0.93]	9.97	80%	
Funding Source							0%	0.11
Government	35	-0.06	0.07	0.005	[-0.19, 0.08]	108.80*	69%	
Private	4	-0.03	0.25	0.064	[-0.52, 0.47]	5.71	47%	
Other/Not Specified	13	-0.01	0.13	0.016	[-0.26, 0.23]	41.66*	71%	
Country of Study							85%	6.87
USA	47	-0.10	0.06	0.003	[-0.21, 0.02]	136.64*	66%	
Great Britain	5	+0.36	0.16	0.027	[0.04, 0.69]	1.68	0%	
Location of Study							95%	40.52*
Urban	5	+0.92	0.16	0.027	[0.60, 1.24]	39.27*	90%	
Suburban	5	-0.36	0.16	0.026	[-0.68, -0.04]	3.19	0%	
Rural	-	-	-	-	-	-	-	
Mixed/Not Specified	42	-0.12	0.05	0.003	[-0.22, -0.02]	57.99	29%	
Experimental Design							0%	0.03
RCT	36	-0.05	0.07	0.013	[-0.18, 0.08]	112.75*	69%	
Quasi-Experimental	16	-0.03	0.11	0.005	[-0.25, 0.19]	43.55*	66%	

Note: All effects included in analysis.

*$p<0.005$

Table E4

Dataset A: Categorical Moderator Analysis of Participant Characteristics (k = 263)

Descriptor	*k*	*g*	SE	Var	95% CI	Q_w	I^2	Q_B
Total Sample Size							86%	7.22
Large (>250)	35	-0.19	0.16	0.026	[-0.50, 0.13]	8478.76*	100%	
Small (<250)	228	+0.28	0.07	0.004	[0.15, 0.41]	691.91*	62%	
School Level							0%	2.57
Preschool-Elementary	225	+0.24	0.07	0.005	[0.11, 0.37]	9173.05*	98%	
Middle School	25	+0.13	0.20	0.040	[-0.26, 0.53]	76.88*	69%	
High School	1	+0.58	1.00	1.003	[-1.38, 2.55]	0.00	0%	
Mixed/Not specified	12	-0.22	0.30	0.087	[-0.80, 0.46]	14.46	24%	
At-Risk in Reading							0%	0.86
0-33%	12	+0.29	0.29	0.081	[-0.27, 0.85]	65.07*	83%	
34-66%	10	+0.11	0.31	0.096	[-0.50, 0.72]	16.85	47%	
67-100%	111	+0.15	0.10	0.009	[-0.04, 0.34]	8410.24*	99%	
Not Specified	130	+0.26	0.09	0.008	[0.09, 0.43]	418.25*	69%	
Low SES							58%	7.20
0-33%	12	+0.42	0.28	0.081	[-0.14, 0.98]	102.95*	89%	
34-66%	63	+0.13	0.13	0.016	[-0.12, 0.38]	126.14*	51%	
67-100%	48	+0.54	0.15	0.021	[0.25, 0.82]	273.32*	83%	
Not Specified	140	+0.12	0.08	0.007	[-0.05, 0.28]	8614.78*	98%	

Note: All effects used in this analysis; SES = socio-economic status as reported in study.
*$p<0.005$

Table E5

Dataset B: Categorical Moderator Analysis of Participant Characteristics (k = 217)

Descriptor	*k*	*g*	SE	Var	95% CI	Q_w	I^2	Q_B
Total Sample Size							91%	11.56*
Large (>250)	122	+0.16	0.02	0.000	[0.12, 0.20]	698.99*	83%	
Small (<250)	95	+0.03	0.03	0.001	[-0.04, 0.10]	147.92*	36%	
School Level							81%	10.68
Preschool-Elementary	79	+0.10	0.03	0.001	[0.05, 0.15]	234.27*	67%	
Middle School	98	+0.19	0.02	0.001	[0.14, 0.24]	377.57*	74%	
High School	40	+0.07	0.07	0.001	[0.01, 0.14]	160.19*	76%	
Mixed/Not specified	-	-	-	-	-	-	-	
At-Risk in Reading							32%	4.41
0-33%	12	+0.09	0.06	0.003	[-0.03, 0.20]	26.27	58%	
34-66%	5	+0.15	0.09	0.008	[-0.03, 0.33]	6.93	42%	
67-100%	38	+0.03	0.06	0.003	[-0.08, 0.14]	62.72	41%	
Not Specified	162	+0.14	0.02	0.000	[0.11, 0.18]	759.41*	79%	
Low SES							30%	4.31
0-33%	1	+0.56	0.28	0.077	[0.02, 1.10]	0.00	0%	
34-66%	18	+0.09	0.05	0.002	[-0.00, 0.19]	34.29	50%	
67-100%	3	+0.28	0.13	0.016	[0.03, 0.52]	4.10	51%	
Not Specified	195	+0.13	0.02	0.000	[0.10, 0.17]	816.88*	76%	

Note: All effects included in analysis. SES = socio-economic status as reported in study.
*$p<0.005$

Table E6

Dataset C: Categorical Moderator Analysis of Participant Characteristics (k = 52)

Descriptor	*k*	*g*	SE	Var	95% CI	Q_w	I^2	Q_B
Total Sample Size							27%	1.37
Large (>250)	7	-0.17	0.12	0.015	[-0.42, 0.07]	8.82	32%	
Small (<250)	45	-0.10	0.07	0.004	[-0.14, 0.12]	141.65*	69%	
School Level							0%	0.05
Preschool-Elementary	46	-0.05	0.06	0.004	[-0.17, 0.07]	151.74*	70%	
Middle School	5	-0.03	0.24	0.058	[-0.50, 0.45]	7.13	44%	
High School	1	+0.04	0.41	0.168	[-0.77, 0.84]	0.00	0%	
Mixed/Not specified	-	-	-	-	-	-	-	
At-Risk in Reading							0%	0.04
0-33%	-	-	-	-	-	-	-	
34-66%	-	-	-	-	-	-	-	
67-100%	38	-0.05	0.07	0.005	[-0.20, 0.09]	128.73*	71%	
Not Specified	14	-0.03	0.10	0.009	[-0.22, 0.16]	29.89	57%	
Low SES							26%	2.72
0-33%	-	-	-	-	-	-	-	
34-66%	10	-0.23	0.14	0.021	[-0.52, 0.05]	9.12	1%	
67-100%	9	+0.08	0.12	0.014	[-0.16, 0.31]	74.93*	89%	
Not Specified	33	-0.04	0.08	0.006	[-0.19, 0.11]	69.15*	54%	

Note: All effects included in analysis. SES = socio-economic status as reported in study.

*$p<0.005$

Table E7

Dataset A: Categorical Moderator Analysis of Intervention Characteristics (k = 263)

Descriptor	*k*	*g*	SE	Var	95% CI	Q_w	I^2	Q_B
Location of Intervention							0%	3.67
Separate Room	139	+0.31	0.09	0.007	[0.14, 0.47]	518.47*	73%	
Core Literacy Classroom	71	+0.03	0.12	0.014	[-0.20, 0.26]	8100.47*	99%	
Mixed	11	+0.20	0.29	0.086	[-0.37, 0.78]	25.71*	61%	
Home	8	+0.22	0.35	0.122	[-0.47, 0.90]	4041	0%	
Other/Not Specified	34	+0.19	0.17	0.029	[-0.14, 0.52]	12.36*	73%	
Literacy Component							0%	1.65
Supplemental	92	+0.24	0.10	0.011	[0.04, 0.44]	346.57*	74%	
Integrated	88	+0.11	0.11	0.011	[-0.10, 0.31]	8250.29*	99%	
Mixed/Not Specified	83	+0.29	0.11	0.012	[0.08, 0.51]	314.24*	74%	
Session Length							0%	0.70
1-29 min	159	+0.25	0.08	0.006	[0.09, 0.41]	8697.45*	98%	
30-59 min	50	+0.15	0.14	0.020	[-0.13, 0.43]	134.69*	64%	
60+ min	39	+0.12	0.16	0.026	[-0.19, 0.44]	149.67*	75%	
Not Specified	15	+0.23	0.25	0.064	[-0.27, 0.73]	89.49*	84%	
Sessions Per Week							63%	8.02
1-2	54	+0.29	0.14	0.018	[0.04, 0.55]	135.84*	61%	
3-4	44	+0.19	0.15	0.023	[-0.11, 0.48]	106.62*	60%	
5	101	+0.03	0.10	0.010	[-0.17, 0.22]	8224.04*	99%	
Not Specified	64	+0.47	0.13	0.016	[0.22, 0.72]	327.13*	81%	
Total Duration							62%	10.50
0-10 hrs	70	+0.43	0.12	0.014	[0.20, 0.66]	207.05*	67%	
11-20 hrs	93	+0.30	0.10	0.010	[0.10, 0.50]	1659.18*	94%	
21-50 hrs	75	-0.00	0.11	0.012	[-0.22, 0.22]	6261.77*	99%	
51-100 hrs	13	-0.11	0.27	0.071	[-0.64, 0.41]	9.40	0%	
Not Specified	12	-0.05	0.27	0.074	[-0.58, 0.48]	23.96	54%	
Teacher							57%	6.92
Licensed	134	+0.17	0.09	0.007	[0.00, 0.33]	8470.18*	98%	
Researchers	48	+0.55	0.15	0.021	[0.26, 0.84]	159.22*	70%	
Other School Staff	33	+0.08	0.17	0.029	[-0.26, 0.41]	125.28*	74%	
Not Specified	48	+0.10	0.14	0.020	[-0.18, 0.38]	155.00*	70%	

Note: All effects used in this analysis.
*$p<0.005$

Table E8

Dataset B: Categorical Moderator Analysis of Intervention Characteristics (k = 217)

Descriptor	*k*	*g*	SE	Var	95% CI	Q_w	I^2	Q_B
Location of Intervention							69%	9.74
Separate Room	18	+0.09	0.05	0.002	[0.00, 0.19]	34.09	50%	
Core Literacy Classroom	4	+0.32	0.12	0.013	[0.10, 0.55]	5.75	48%	
Mixed	9	-0.10	0.09	0.009	[-0.28, 0.08]	6.71	0%	
Home	-	-	-	-	-	-	-	
Other/Not Specified	186	+0.14	0.02	0.000	[0.10, 0.17]	798.12*	77%	
Literacy Component							68%	6.20
Supplemental	9	-0.10	0.09	0.009	[-0.28, 0.08]	6.71	0%	
Integrated	42	+0.13	0.04	0.002	[0.05, 0.21]	62.89	35%	
Mixed/Not Specified	166	+0.14	0.02	0.000	[0.10, 0.17]	784.74*	79%	
Session Length							72%	10.56
1-29 min	4	+0.12	0.12	0.013	[0.10, 0.55]	5.75	48%	
30-59 min	187	+0.13	0.02	0.000	[0.10, 0.16]	817.89*	77%	
60+ min	20	+0.15	0.10	0.009	[-0.03, 0.34]	13.37	0%	
Not Specified	6	-0.30	0.15	0.024	[-0.56, 0.01]	14.83	66%	
Sessions Per Week							70%	10.01
1-2	16	+0.09	0.05	0.002	[-0.01, 0.19]	30.80	51%	
3-4	171	+0.14	0.02	0.000	[0.10, 0.17]	778.31*	78%	
5	24	+0.22	0.07	0.005	[0.08, 0.37]	21.05	0%	
Not Specified	6	-0.30	0.15	0.024	[-0.60, 0.01]	14.83	66%	
Total Duration							85%	25.95*
0-10 hrs	89	+0.06	0.02	0.001	[0.02, 0.11]	368.04*	76%	
11-20 hrs	98	+0.19	0.02	0.001	[0.15, 0.24]	408.84*	76%	
21-50 hrs	4	+0.32	0.11	0.013	[0.10, 0.55]	5.75	48%	
51-100 hrs	20	+0.15	0.09	0.009	[-0.03, 0.34]	13.37	0%	
Not Specified	6	-0.30	0.15	0.023	[-0.60, 0.00]	14.83	66%	
Teacher							68%	6.24
Licensed	22	+0.13	0.04	0.002	[0.04, 0.21]	49.21	57%	
Researchers	9	-0.10	0.09	0.009	[-0.28, 0.08]	6.71	0%	
Other School Staff	-	-	-	-	-	-	-	
Not Specified	186	+0.14	0.02	0.000	[0.10, 0.17]	798.12*	77%	

Note: All effects used in this analysis

*$p<0.005$

Table E9

Dataset C: Categorical Moderator Analysis of Intervention Characteristics (k = 52)

Descriptor	*k*	*g*	SE	Var	95% CI	Q_w	I^2	Q_B
Location of Intervention							67%	9.11
Separate Room	36	-0.08	0.07	0.005	[-0.21, 0.06]	105.70*	67%	
Core Literacy Classroom	9	-0.11	0.11	0.011	[-0.32, 0.10]	14.53	45%	
Mixed	5	-0.03	0.22	0.054	[-0.48, 0.43]	7.13	44%	
Home	-	-	-	-	-	-	-	
Other/Not Specified	2	+0.71	0.26	0.068	[0.20, 1.22]	5.27	81%	
Literacy Component							57%	4.71
Supplemental	26	+0.08	0.08	0.007	[-0.09, 0.24]	117.31*	79%	
Integrated	14	-0.10	0.10	0.010	[-0.29, 0.10]	22.09	41%	
Mixed/Not Specified	12	-0.22	0.16	0.013	[-0.45, 0.01]	3.25	0%	
Session Length							73%	11.25
1-29 min	16	-0.10	0.08	0.007	[-0.27, 0.07]	35.97*	58%	
30-59 min	3	+0.66	0.22	0.047	[0.23, 1.08]	5.56	64%	
60+ min	28	-0.10	0.08	0.006	[-0.25, 0.06]	81.71*	67%	
Not Specified	5	-0.03	0.23	0.054	[-0.48, 0.43]	7.13	44%	
Sessions Per Week							93%	44.87*
1-2	8	+0.55	0.12	0.014	[0.32, 0.78]	42.78*	84%	
3-4	7	+0.34	0.15	0.023	[0.04, 0.64]	20.56*	71%	
5	32	-0.21	0.56	0.003	[-0.31, -0.10]	23.25	0%	
Not Specified	5	-0.36	0.16	0.024	[-0.67, -0.05]	3.19	0%	
Total Duration							74%	11.36
0-10 hrs	12	+0.23	0.12	0.013	[0.04, 0.49]	74.32*	85%	
11-20 hrs	6	-0.01	0.20	0.040	[-0.40, 0.38]	7.17	30%	
21-50 hrs	22	-0.20	0.08	0.006	[-0.36, -0.05]	13.89	0%	
51-100 hrs	12	-0.03	0.12	0.015	[-0.26, 0.21]	37.84*	71%	
Not Specified	-	-	-	-	-	-	-	
Teacher							95%	36.80*
Licensed	41	-0.19	0.05	0.003	[-0.29, -0.09]	40.54	1%	
Researchers	9	+0.47	0.12	0.015	[0.23, 0.71]	50.51*	84%	
Other School Staff	-	-	-	-	-	-	-	
Not Specified	2	+0.71	0.22	0.049	[0.28, 1.14]	5.27	81%	

Note: All effects included in analysis.
*$p<0.005$

Table E10

Dataset A: Categorical Moderator Analysis of CAI Characteristics (k = 263)

Descriptor	*k*	*g*	SE	Var	95% CI	Q_w	I^2	Q_B
CAI							0%	0.03
Commercially Available	225	+0.21	0.07	0.004	[0.08, 0.34]	9116.11*	98%	
Not Commercially Available	38	+0.16	0.16	0.026	[-0.18, 0.55]	128.64*	71%	
CAI Structure							26%	2.72
Linear	168	+0.14	0.08	0.006	[-0.02, 0.28]	8729.35*	98%	
Adaptive	62	+0.34	0.13	0.016	[0.09, 0.59]	98.30*	38%	
Intervention Components								
Guided Activity							50%	2.01
Yes	85	+0.34	0.11	0.012	[0.13, 0.55]	215.39*	61%	
No/Not Reported	178	+0.15	0.07	0.005	[0.01, 0.30]	8955.49*	98%	
Corrective Feedback							45%	1.81
Yes	204	+0.21	0.07	0.005	[0.07, 0.34]	8949.68*	98%	
No/Not Reported	59	+0.22	0.13	0.016	[-0.03, 0.47]	229.93*	75%	
Elaborate Feedback							0%	0.37
Yes	57	+0.14	0.13	0.018	[-0.12, 0.40]	111.32*	50%	
No/Not Reported	206	+0.23	0.07	0.005	[0.09, 0.37]	9149.91*	98%	
Pacing							53%	2.13
Yes	174	+0.27	0.08	0.006	[0.13, 0.42]	510.06*	66%	
No/Not Reported	89	+0.09	0.11	0.011	[-0.12, 0.29]	8636.24*	99%	
Pretraining							67%	3.10
Yes	14	-0.24	0.26	0.069	[-0.75, 0.28]	21.10	38%	
No/Not Reported	249	+0.24	0.06	0.004	[0.11, 0.36]	9246.30*	97%	
Record Keeping							0%	0.01
Yes	211	+0.21	0.07	0.005	[0.07, 0.34]	9030.51*	98%	
No/Not Reported	52	+0.22	0.14	0.019	[-0.05, 0.49]	101.44*	50%	
Number of Components							27%	2.73
0-1 Components	45	+0.33	0.15	0.022	[0.04, 0.62]	104.69*	58%	
2-3 Components	123	+0.10	0.09	0.008	[-0.07, 0.28]	8494.26*	99%	
4-6 Components	95	+0.29	0.10	0.010	[0.09, 0.49]	278.18*	66%	

Note: All effects used in this analysis.

**p*<0.005

Table E11

Dataset B: Categorical Moderator Analysis of CAI Characteristics (k = 217)

Descriptor	*k*	*g*	SE	Var	95% CI	Q_w	I^2	Q_B
CAI							0%	0.00
Commercially Available	217	+0.13	0.02	0.000	[0.10, 0.16]	870.63*	75%	
Not Commercially Available	-	-	-	-	-	-	-	
CAI Structure							40%	3.34
Linear	26	+0.11	0.07	0.004	[-0.02, 0.24]	47.01*	47%	
Adaptive	160	+0.15	0.02	0.000	[0.11, 0.18]	756.17*	79%	
Not Specified	31	+0.06	0.04	0.002	[-0.02, 0.15]	53.71	44%	
Intervention Components								
Guided Activity							0%	0.00
Yes	-	-	-	-	-	-	-	
No/Not Reported	217	+0.13	0.02	0.000	[0.10, 0.16]	870.63*	75%	
Corrective Feedback							66%	2.92
Yes	170	+0.14	0.02	0.000	[0.11, 0.18]	795.43*	79%	
No/Not Reported	47	+0.07	0.04	0.001	[-0.01, 0.15]	61.16	25%	
Elaborate Feedback							0%	0.00
Yes	-	-	-	-	-	-	-	
No/Not Reported	217	+0.13	0.02	0.000	[0.10, 0.16]	870.63*	75%	
Pacing							83%	5.86
Yes	164	+0.15	0.02	0.000	[0.11, 0.19]	766.75*	79%	
No/Not Reported	53	+0.05	0.04	0.001	[-0.03, 0.12]	85.37*	39%	
Pretraining							0%	0.00
Yes	-	-	-	-	-	-	-	
No/Not Reported	217	+0.13	0.02	0.000	[0.10, 0.16]	870.63*	75%	
Record Keeping							92%	13.07*
Yes	202	+0.14	0.02	0.000	[0.11, 0.18]	824.25*	76%	
No/Not Reported	15	-0.15	0.08	0.006	[-0.31, 0.00]	23.71	41%	
Number of Components							83%	5.86
0-1 Components	53	+0.05	0.04	0.001	[-0.03, 0.12]	83.37*	39%	
2-3 Components	164	+0.15	0.02	0.000	[0.11, 0.19]	766.75*	79%	
4-6 Components	-	-	-	-	-	-	-	

Note: All effects used in this analysis

**p*<0.005

Table E12

Dataset C: Categorical Moderator Analysis of CAI Characteristics (k = 52)

Descriptor	*k*	*g*	SE	Var	95% CI	Q_w	I^2	Q_B
CAI							84%	6.37
Commercially Available	46	-0.10	0.06	0.003	[-0.22, 0.01]	136.19*	67%	
Not Commercially Available	6	+0.31	0.15	0.023	[0.01, 0.61]	2.95	0%	
CAI Structure							0%	0.20
Linear	39	-0.06	0.07	0.004	[-0.19, 0.07]	119.03*	68%	
Adaptive	13	+0.00	0.12	0.015	[-0.24, 0.25]	37.21*	68%	
Not Specified	-	-	-	-	-	-	-	
Intervention Components								
Guided Activity							93%	13.81*
Yes	7	+0.65	0.20	0.038	[0.27, 1.03]	49.20*	88%	
No/Not Reported	45	-0.11	0.06	0.003	[-0.22, 0.01]	88.36*	50%	
Corrective Feedback							64%	2.75
Yes	43	-0.09	0.06	0.004	[-0.21, 0.04]	84.55*	50%	
No/Not Reported	9	+0.19	0.15	0.023	[-0.11, 0.48]	70.23*	87%	
Elaborate Feedback							0%	0.01
Yes	5	-0.03	0.24	0.058	[-0.50, 0.45]	7.13	44%	
No/Not Reported	47	-0.05	0.06	0.004	[-0.16, 0.07]	152.01*	70%	
Pacing							0%	0.25
Yes	24	-0.02	0.08	0.007	[-0.18, 0.14]	73.48*	69%	
No/Not Reported	28	-0.08	0.09	0.007	[-0.24, 0.09]	85.71*	68%	
Pretraining							98%	49.45*
Yes	2	+2.26	0.33	0.111	[1.61, 2.91]	1.10	9%	
No/Not Reported	50	-0.10	0.05	0.002	[-0.19, -0.01]	95.70*	49%	
Record Keeping							87%	7.81
Yes	38	-0.14	0.07	0.004	[-0.27, -0.01]	64.56*	43%	
No/Not Reported	14	+0.21	0.11	0.011	[-0.00, 0.42]	78.09*	83%	
Number of Components							0%	1.51
0-1 Components	8	-0.22	0.15	0.023	[-0.52, 0.08]	10.49	33%	
2-3 Components	39	-0.02	0.07	0.004	[-0.14, 0.11]	140.32*	73%	
4-6 Components	5	-0.03	0.24	0.058	[-0.50, 0.45]	7.13	44%	

Note: All effects included in analysis.
*$p<0.005$

Appendix F
Summary of Quality Indicators Met by Each Study

Table F1

Summary of Quality Indicators (Gersten et al., 2005)

			Participant Description				Intervention Implementation			Outcome Measures		Data Analysis				
Study Citation	Pub Type	Data Set	Abil/Dis Des	Comp Grps	Comp Int	Att < 30%	Int Des	Fid	Cont Des	Mult Meas	Meas Timed	Data Anal	Effect Sizes	RCT	Score	%
Arvans (2010)	DI	A	1	1	0	1	1	0	1	1	1	1	1	1	10	83%
Beaird (2007)	DI	A	0	1	0	0	1	0	1	0	1	1	0	1	6	50%
Bjorn & Leppanen (2013)	JN	A	1	0	0	0	1	0	0	0	1	1	0	1	5	42%
Brooks et al. (2006)	JN	A	0	0	0	1	0	0	0	0	1	1	0	1	4	33%
Campuzano et al. (2009)	TR	A	0	1	0	0	1	1	0	0	1	1	1	1	7	58%
Cassady & Smith (2004)	JN	A	0	0	0	0	1	0	0	0	1	1	0	0	3	25%
Cassady & Smith (2005)	JN	B	1	1	1	0	1	0	1	0	1	1	0	0	7	58%
Chera & Wood (2003)	JN	A	0	1	0	0	1	0	0	0	1	0	0	0	3	25%
Christ & Davie (2009)	TR	A	1	1	1	1	1	1	1	0	1	1	1	1	11	92%
Cohen et al. (2005)	JN	A	1	0	0	0	1	0	0	0	0	1	0	1	4	33%
Cole & Hilliard (2006)	JN	C	1	1	0	0	1	0	1	1	1	1	0	1	8	67%
Cuevas et al. (2012)	JN	AC	0	0	1	1	1	0	1	0	1	1	0	1	7	58%
Fehr (2011)	DI	A	0	1	0	0	1	0	0	0	1	1	1	1	6	50%
Fischel et al. (2007)	JN	AC	1	1	0	0	1	1	1	0	1	1	1	1	9	75%
Gale (2006)	DI	A	1	0	1	0	1	0	0	1	1	1	1	1	8	67%
Gatti & Miller (2011)	TR	B	1	0	1	1	0	0	1	1	1	1	1	1	9	75%
Gillam et al. (2008)	JN	AC	1	1	1	1	1	1	1	0	1	1	1	1	11	92%
Gillard (2010)	DI	A	0	0	0	0	1	0	0	0	0	1	0	0	2	17%
Given et al. (2008)	JN	B	1	1	0	1	1	0	0	0	1	1	1	1	8	67%
Green et al. (2011)	TR	A	0	1	0	0	1	0	0	0	1	0	0	1	4	33%
Greenlee (2001)	DI	B	1	1	0	0	1	0	0	0	1	0	0	0	4	33%
Hansen et al. (2004)	TR	A	1	1	1	0	1	1	1	0	1	1	1	0	9	75%
Hasselbring & Goin (2004)	JN	A	0	0	0	0	1	0	0	0	1	0	0	0	2	17%
Hecht & Close (2002)	JN	A	0	0	0	0	0	0	0	0	1	1	1	1	4	33%
Higgins & Raskind (2004)	JN	A	1	0	0	0	1	0	0	0	0	1	1	1	5	42%
Hook et al. (2001)	JN	AC	0	0	0	0	1	0	0	0	1	1	0	0	3	25%
Huffstetter et al. (2010)	JN	A	1	1	1	0	1	1	1	0	1	1	1	1	10	83%
Johnson et al. (2010)	JN	A	1	1	0	0	1	0	0	1	1	1	0	1	7	58%
Jones et al. (2004)	JN	A	0	0	0	0	1	0	1	1	1	1	0	0	5	42%
Khan & Gorard (2012)	JN	A	0	1	0	1	1	0	0	0	1	1	1	1	7	58%
Kim et al. (2006)	JN	A	1	1	1	0	1	0	0	0	1	0	1	1	7	58%

Table F1 *(cont.)*

Summary of Quality Indicators (Gersten et al., 2005)

			Participant Description				Intervention Implementation			Outcome Measures		Data Analysis				
Study Citation	Pub Type	Data Set	Abil/Dis Des	Comp Grps	Comp Int	Att < 30%	Int Des	Fid	Cont Des	Mult Meas	Meas Timed	Data Anal	Effect Sizes	RCT	Score	%
Kreskey (2012)	DI	A	1	1	0	0	1	0	1	1	1	1	0	0	7	58%
Longberg (2012)	DI	B	1	0	1	1	1	1	1	0	1	1	1	1	10	83%
Lonigan et al. (2003)	JN	A	1	1	1	1	1	0	0	0	1	1	1	1	9	75%
Macaruso et al. (2006)	JN	A	1	1	1	0	1	0	1	0	1	1	0	1	8	67%
Macaruso & Rodman (2009)	JN	A	1	1	1	0	1	0	1	0	1	1	1	1	9	75%
Macaruso & Rodman (2011a)	JN	A	1	1	0	0	1	1	1	1	1	1	1	1	10	83%
Macaruso & Rodman (2011b)	JN	A	0	1	0	0	1	0	0	1	1	1	1	0	6	50%
Macaruso & Walker (2008)	JN	A	0	0	1	1	1	0	0	0	1	1	1	1	7	58%
Mathes et al. (2001)	JN	B	1	0	1	1	0	0	1	1	1	1	1	1	9	75%
Mitchell & Fox (2001)	JN	AC	1	1	1	0	1	0	1	1	1	1	0	1	9	75%
Mostow et al. (2003)	JN	A	0	1	0	0	1	1	1	1	1	1	1	0	8	67%
Nicolson et al. (2000)	JN	AC	1	1	0	0	0	0	1	0	1	1	0	0	5	42%
Paterson et al. (2003)	JN	A	0	0	0	0	1	0	0	1	1	1	0	0	4	33%
Pokorni et al. (2004)	JN	C	1	1	1	1	0	0	1	0	1	1	1	0	8	67%
Powers & Price-Johnson (2006)	TR	A	0	0	0	0	0	0	0	1	1	1	1	0	4	33%
Rasinski et al. (2011)	JN	B	0	0	0	0	1	0	0	1	0	1	1	0	4	33%
Rehmann (2005)	DI	A	0	0	0	0	1	0	1	0	1	1	0	1	5	42%
Reutzel et al. (2012)	JN	C	1	0	0	0	1	0	1	1	0	1	1	0	6	50%
Rouse & Krueger (2004)	JN	A	0	1	0	0	1	0	0	0	1	1	0	1	5	42%
Rupprecht (2003)	DI	C	1	0	0	0	0	0	1	0	1	0	0	1	4	33%
Shelley-Tremblay & Eyer (2009)	JN	A	0	0	1	0	1	0	1	1	1	0	0	1	6	50%
Stevenson (2006)	TR	A	0	1	1	0	0	1	1	0	1	1	0	0	6	50%
Tracey & Young (2007)	JN	B	0	0	0	1	1	0	0	1	1	1	0	1	6	50%
Troia (2004)	JN	A	1	1	0	0	1	1	0	1	1	1	1	0	8	67%
Troia & Whitney (2003)	JN	A	0	1	0	0	1	0	0	0	1	1	1	0	5	42%
Valliath (2002)	DI	A	0	1	0	0	1	0	1	0	1	1	0	1	6	50%
Wijekumar et al. (2013)	JN	A	0	1	0	0	1	1	0	0	1	1	1	1	7	58%
Wijekumar et al. (2012)	JN	A	0	1	0	0	1	1	0	0	1	1	1	1	7	58%

Table F1 *(cont.)*

Summary of Quality Indicators (Gersten et al., 2005)

			Participant Description				Intervention Implementation			Outcome Measures		Data Analysis				
Study Citation	Pub Type	Data Set	Abil/Dis Des	Comp Grps	Comp Int	Att < 30%	Int Des	Fid	Cont Des	Mult Meas	Meas Timed	Data Anal	Effect Sizes	RCT	Score	%
Wild (2009)	JN	AC	0	1	1	0	1	0	1	1	1	1	0	1	8	67%
Woodward (2005)	DI	A	0	0	0	0	1	0	0	0	1	1	0	0	3	25%
Percent of Indicators Met		A	41%	63%	31%	16%	90%	24%	43%	29%	94%	90%	49%	65%		52%
		B	75%	38%	50%	63%	75%	13%	50%	50%	88%	88%	63%	63%		59%
		C	73%	64%	45%	27%	73%	18%	91%	36%	91%	91%	36%	64%		59%
Overall (*n* = 61)			49%	59%	33%	23%	85%	21%	48%	33%	92%	89%	51%	64%		54%

Note: 1 = indicator present, 0 = indicator not present; DI = Dissertation; JN = Journal Article; TR = Technical Report

Appendix G
Categorical Moderator Analyses of Quality Indicators

Table G1

Dataset A: Categorical Analysis of Research Quality (k = 66)

Descriptor	*k*	*g*	SE	Var	95% CI	Q_w	I^2	Q_B
Participant Description								
Participant Abilities Described							0%	0.28
Yes	27	+0.29	0.11	0.012	[0.07, 0.50]	877.41*	97%	
No	39	+0.21	0.09	0.008	[0.04, 0.39]	202.92*	81%	
Comparable Groups							0%	0.28
Yes	42	+0.22	0.09	0.007	[0.05, 0.39]	1005.20*	96%	
No	24	+0.29	0.12	0.014	[0.06, 0.53]	66.71*	66%	
Comparable Interventionists							60%	2.50
Yes	20	+0.41	0.13	0.016	[0.16, 0.66]	866.30*	98%	
No	46	+0.17	0.08	0.007	[0.01, 0.33]	214.45*	79%	
Attrition (<30%)							0%	0.27
Yes	9	+0.16	0.18	0.033	[-0.20, 0.51]	28.10*	72%	
No	57	+0.26	0.08	0.006	[0.11, 0.40]	1047.49*	97%	
Intervention/Control Description								
Adequate Intervention Description							0%	0.21
Yes	61	+0.23	0.07	0.005	[0.09, 0.38]	1048.17*	94%	
No	5	+0.35	0.24	0.059	[-0.13, 0.82]	13.68	71%	
Fidelity of Implementation							0%	0.74
Yes	20	+0.16	0.12	0.015	[-0.08, 0.40]	863.83*	98%	
No	46	+0.29	0.09	0.008	[0.12, 0.46]	207.36*	78%	
Adequate Control Description							0%	0.10
Yes	25	+0.21	0.11	0.013	[-0.01, 0.43]	862.92*	97%	
No	41	+0.26	0.09	0.008	[0.09, 0.43]	217.91*	82%	
Outcome Measures								
Multiple Measures							57%	2.30
Yes	19	+0.41	0.13	0.017	[0.15, 0.67]	58.73*	69%	
No	47	+0.12	0.08	0.007	[0.02, 0.34]	995.17*	95%	
Measures Timed Appropriately							0%	0.49
Yes	62	+0.26	0.07	0.005	[0.11, 0.40]	1078.06*	94%	
No	4	+0.29	0.08	0.081	[-0.51, 0.61]	1.93	0%	
Data Analysis								
Appropriate to Unit of Analysis							0%	0.00
Yes	61	+0.25	0.08	0.006	[0.09, 0.40]	1071.74*	94%	
No	5	+0.24	0.27	0.072	[-0.28, 0.77]	7.61	47%	
Effect Sizes Reported							0%	0.32
Yes	37	+0.28	0.10	0.009	[0.09, 0.47]	949.40*	96%	
No								
Research Design								
Randomized Control Trial							16%	1.19
Yes	46	+0.29	0.09	0.007	[0.13, 0.46]	235.70*	81%	
No	20	+0.13	0.13	0.016	[-0.12, 0.38]	844.70*	98%	
Quality Score							0%	0.99
0-25%	6	+0.17	0.24	0.058	[-0.30, 0.65]	6.36	21%	
26-50%	25	+0.26	0.12	0.015	[0.02, 0.50]	122.58*	80%	
51-75%	29	+0.21	0.11	0.012	[-0.01, 0.42]	897.55*	97%	
76-100%	6	+0.45	0.24	0.057	[-0.01, 0.92]	18.09*	72%	

Note: Only independent effects included in analysis.

**p*<0.005

Table G2

Dataset B: Categorical Analysis of Research Quality (k = 22)

Descriptor	*k*	*g*	SE	Var	95% CI	Q_w	I^2	Q_B
Participant Description								
Participant Abilities Described							0%	0.49
Yes	14	+0.12	0.06	0.004	[-0.00, 0.23]	27.66	53%	
No	8	+0.17	0.05	0.002	[0.08, 0.26]	20.49	66%	
Comparable Groups							0%	0.46
Yes	9	+0.07	0.12	0.015	[-0.17, 0.31]	23.70*	66%	
No	13	+0.16	0.04	0.001	[0.08, 0.23]	24.67	51%	
Comparable Interventionists							0%	0.07
Yes	6	+0.16	0.07	0.005	[0.03, 0.30]	7.72	35%	
No	16	+0.14	0.04	0.002	[0.06, 0.23]	40.97*	63%	
Attrition (<30%)							0%	0.00
Yes	9	+0.15	0.06	0.004	[0.03, 0.27]	5.18	0%	
No	13	+0.15	0.05	0.002	[0.06, 0.24]	43.84*	72%	
Intervention/Control Description								
Adequate Intervention Description							0%	0.07
Yes	18	+0.16	0.04	0.002	[0.07, 0.24]	44.67*	62%	
No	4	+0.13	0.08	0.006	[-0.02, 0.28]	3.83	22%	
Fidelity of Implementation							0%	0.03
Yes	1	+0.12	0.17	0.027	[-0.20, 0.45]	0.00	0%	
No	21	+0.15	0.04	0.001	[0.08, 0.22]	48.62*	59%	
Adequate Control Description							0%	0.07
Yes	6	+0.16	0.07	0.005	[0.03, 0.30]	7.72	35%	
No	16	+0.14	0.04	0.002	[0.06, 0.23]	40.97*	63%	
Outcome Measures								
Multiple Measures							0%	0.44
Yes	12	+0.16	0.04	0.002	[0.08, 0.24]	24.58	55%	
No	10	+0.09	0.10	0.010	[-0.10, 0.28]	23.74	62%	
Measures Timed Appropriately							0%	0.13
Yes	15	+0.13	0.06	0.003	[0.02, 0.25]	28.87	52%	
No	7	+0.16	0.05	0.002	[0.07, 0.26]	19.69*	70%	
Data Analysis								
Appropriate to Unit of Analysis							81%	5.16
Yes	17	+0.17	0.03	0.001	[0.10, 0.23]	28.52	44%	
No	5	-0.26	0.18	0.033	[-0.61, 0.10]	14.63	73%	
Effect Sizes Reported							0%	0.00
Yes	15	+0.15	0.04	0.001	[0.07, 0.23]	24.04	42%	
No	7	+0.15	0.11	0.012	[-0.07, 0.36]	24.56*	76%	
Research Design								
Randomized Control Trial							0%	0.00
Yes	9	+0.15	0.06	0.004	[0.03, 0.27]	5.18	0%	
No	13	+0.15	0.05	0.002	[0.06, 0.24]	43.84*	72%	
Quality Score							0%	0.11
0-25%	-	-	-	-	-	-	-	
26-50%	13	+0.14	0.05	0.002	[0.05, 0.23]	40.72*	71%	
51-75%	8	+0.17	0.07	0.005	[0.03, 0.31]	7.90	11%	
76-100%	1	+0.12	0.17	0.028	[-0.21, 0.46]	0.00	0%	

Note: Only independent effects included in analysis.

*p<0.005

Table G3

Dataset C: Categorical Analysis of Research Quality (k = 13)

Descriptor	*k*	*g*	SE	Var	95% CI	Q_w	I^2	Q_B
Participant Description								
Participant Abilities Described							0%	0.08
Yes	10	+0.16	0.17	0.029	[-0.18, 0.50]	49.86*	82%	
No	3	+0.06	0.32	0.101	[0.56, 0.69]	1.90	0%	
Comparable Groups							0%	0.02
Yes	9	+0.13	0.17	0.030	[-0.21, 0.46]	42.08*	81%	
No	4	+0.17	0.28	0.077	[-0.38, 0.71]	6.85	56%	
Comparable Interventionists							68%	3.20
Yes	7	-0.10	0.21	0.042	[-0.51, 0.30]	7.01	14%	
No	6	+0.46	0.24	0.055	[-0.00, 0.92]	43.37*	88%	
Attrition (<30%)							51%	2.05
Yes	5	-0.13	0.24	0.057	[-0.59, 0.34]	1.73	0%	
No	8	+0.32	0.20	0.038	[-0.07, 0.70]	48.19*	85%	
Intervention/Control Description								
Adequate Intervention Description							0%	0.61
Yes	9	+0.18	0.18	0.031	[-0.16, 0.53]	47.56*	83%	
No	4	+0.02	0.29	0.082	[-0.55, 0.58]	4.86	38%	
Fidelity of Implementation							33%	1.49
Yes	3	-0.14	0.26	0.069	[-0.65, 0.39]	0.28	0%	
No	10	+0.25	0.17	0.028	[-0.08, 0.57]	42.42*	79%	
Adequate Control Description							0%	0.54
Yes	12	+0.16	0.15	0.023	[-0.13, 0.46]	51.84*	79%	
No	1	-0.31	0.62	0.388	[-1.53, 0.91]	0.00	0%	
Outcome Measures								
Multiple Measures							81%	5.34
Yes	4	+0.58	0.23	0.055	[0.12, 1.04]	30.48*	90%	
No	9	-0.07	0.16	0.024	[-0.38, 0.24]	6.64	0%	
Measures Timed Appropriately							43%	1.75
Yes	12	+0.08	0.14	0.021	[-0.21, 0.36]	42.55*	74%	
No	1	+0.72	0.47	0.218	[-0.19, 1.64]	0.00	0%	
Data Analysis								
Appropriate to Unit of Analysis							0%	0.07
Yes	12	+0.15	0.15	0.023	[-0.15, 0.44]	51.42*	79%	
No	1	-0.02	0.63	0.391	[-1.25, 1.20]	0.00	0%	
Effect Sizes Reported							15%	1.18
Yes	6	-0.02	0.20	0.041	[-0.42, 0.38]	14.48	65%	
No	7	+0.23	0.21	0.043	[-0.11, 0.71]	32.70*	82%	
Research Design								
Randomized Control Trial							0%	0.00
Yes	8	+0.13	0.18	0.033	[-0.23, 0.49]	37.15*	81%	
No	5	+0.14	0.25	0.061	[-0.34, 0.63]	11.91	66%	
Quality Score							0%	2.00
0-25%	1	-0.31	0.63	0.397	[-1.54, 0.93]	0.00	0%	
26-50%	3	+0.46	0.32	0.103	[-0.17, 1.09]	2.30	13%	
51-75%	7	+0.14	0.20	0.041	[-0.26, 0.54]	38.35*	84%	
76-100%	2	-0.11	0.35	0.122	[-0.80, 0.57]	0.16	0%	

Note: Only independent effects included in analysis.
$*p<0.005$

Appendix H
Reference List of Studies Included in Meta-Analyses

Arvans, R. (2010). Improving reading fluency and comprehension in elementary students using read naturally. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 71*(1-B), 649.

Beaird, C. K. (2007). The effects of computer-assisted language learning on english language learners with and without disabilities in an elementary school setting. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 68*(4-A), 1421.

Björn, P. M., & Leppänen, P. H. T. (2013). Accelerating decoding-related skills in poor readers learning a foreign language: A computer-based intervention. *Educational Psychology, 33*(6), 671-689. doi:10.1080/01443410.2013.797336

Brooks, G., Miles, J., Torgerson, C. J., & Torgerson, D. J. (2006). Is an intervention using computer software effective in literacy learning? A randomised controlled trial. *Educational Studies, 32*(2), 133-143. doi:10.1080/03055690500416116

Campuzano, L., Dynarski, M., Agodini, R., Rall, K., & Institute of Education Sciences, National Center for Education Evaluation and, Regional Assistance. (2009). *Effectiveness of reading and mathematics software products: Findings from two student cohorts. NCEE 2009-4041.*National Center for Education Evaluation and Regional Assistance.

Cassady, J., & Smith, L. (2005). The impact of a structured integrated learning system on first-grade students' reading gains. *Reading & Writing Quarterly, 21*(4), 361-376.

Cassady, J., & Smith, L. (2004). The impact of a reading- focused integrated learning system on phonological awareness in kindergarten. *Journal of Literacy Research, 35*(4), 947-964. doi:10.1207/s15548430jlr3504_2

Chera, P., & Wood, C. (2003). Animated multimedia ‘talking books’ can promote phonological awareness in children beginning to read. *Learning and Instruction, 13*(1), 33-52. doi: 10.1016/S0959-4752(01)00035-4

Christ, T., & Davie, J. (2009). Empirical evaluation of Read Naturally® effects: A randomized control trial. University of Minnesota.

Cohen, W., Hodson, A., O'Hare, A., Boyle, J., Durrani, T., McCartney, E., et al. (2005). Effects of computer-based intervention through acoustically modified speech (fast ForWord) in severe mixed receptive-expressive language impairment: Outcomes from a randomized controlled trial. *Journal of Speech, Language & Hearing Research, 48*(3), 715-729.

Cole, J. M., & Hilliard, V. R. (2006). The effects of web-based reading curriculum on children's reading performance and motivation.*Journal of Educational Computing Research, 34*(4), 353-380.

Cuevas, J. A., Russell, R. L., & Irving, M. A. (Jun 2012). An examination of the effect of customized reading modules on diverse secondary students' reading comprehension and motivation. *Educational Technology Research and Development, 60*(3), 445-467.

Fehr, C. N. (2012). Analysis of two randomized field trials testing the effects of online vocabulary instruction on vocabulary test scores. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 73*(2-A), 557.

Fischel, J. E., Bracken, S. S., Fuchs-Eisenberg, A., Spira, E. G., Katz, S., & Shaller, G. (2007). Evaluation of curricular approaches to enhance preschool early literacy skills. *Journal of Literacy Research, 39*(4), 471-501. doi:10.1080/10862960701675333

Gale, Deanne (2006) The Effect of Computer-Delivered Phonological Awareness Training on the Early Literacy Skills of Students Identified as At-Risk for Reading Failure (Doctoral dissertation). *Graduate School Theses and Dissertations.*

Gatti, G. G. & Miller, J. (2011). *SuccessMaker: Pearson SuccessMaker reading efficacy study. 2010-11 Final Report*. Pittsburgh, PA: Gatti Evaluation Inc.

Gillam, R. B., Loeb, D. F., Hoffman, L. M., Bohman, T., Champlin, C. A., Thibodeau, L., et al. (2008). The efficacy of fast ForWord language intervention in school-age children with language impairment: A randomized controlled trial. *Journal of Speech, Language & Hearing Research, 51*(1), 97-119.

Gillard, D. J. (2010). An investigation of the impact of computer assisted instruction on reading comprehension of middle school sixth grade students in a rural south Carolina school district. (Ed.D., South Carolina State University). *ProQuest Dissertations and Theses.*

Given, B. K., Wasserman, J. D., Chari, S. A., Beattie, K., & Eden, G. F. (Aug 2008). A randomized, controlled study of computer-based intervention in middle school struggling readers. *Brain and Language, 106*(2), 83-97

Green, P. J., Sha, M., & Liu, L. (2011) U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service, *The U.S.-China E-Language Project: A Study of a Gaming Approach to English Language Learning for Middle School Students.* Washington, D.C., 2011.

Greenlee, N. B. (2001). An analysis of the effectiveness of the "academy of reading" training on the reading achievement of elementary students in a suburban, Midwestern school district. (Doctoral dissertation). *ProQuest Dissertations and Theses.*

Hansen, E. E., Llosa, L., & Slayton, J. (2004). *Evaluation of the Waterford Early Reading Program as a supplementary program in the Los Angeles Unified School District: 2002-03.* (Publication No. 177).

Hasselbring, T. S., & Goin, L. I. (2004). Literacy instruction for older struggling readers: What is the role of technology? *Reading and Writing Quarterly, 20*(2), 123-144.

Hecht, S. A., & Close, L. (2002). Emergent literary skills and training time uniquely predict variability in responses to phonemic awareness training in disadvantaged kindergartners. *Journal of Experimental Child Psychology, 82*(2), 93-115

Higgins, E. L., & Raskind, M. H. (2004). Speech recognition-based and automaticity programs to help students with severe reading and spelling problems. *Annals of Dyslexia, 54*(2), 365-392.

Hook, P. E., Macaruso, P., & Jones, S. (2001). Efficacy of Fast ForWord training on facilitating acquisition of reading skills by children with reading difficulties – A longitudinal study. *Annals of Dyslexia, 51*, 75-96

Huffstetter, M., King, J. R., Onwuegbuzie, A. J., Schneider, J. J., & Powell-Smith, K. (2010). Effects of a computer-based early reading program on the early reading and oral language skills of at-risk preschool children. *Journal of Education for Students Placed at Risk (JESPAR), 15*(4), 279-298. doi:10.1080/10824669.2010.532415

Johnson, E. P., Perry, J., & Shamir, H. (Aug 2010). Variability in reading ability gains as a function of computer-assisted instruction method of presentation. *Computers & Education, 55*(1), 209-217.

Jones, J. D., Staats, W. D., Bowling, N., Bickel, R. D., Cunningham, M. L., & Cadle, C. (2004). An evaluation of the merit reading software program in the Calhoun county (WV) Middle/High school. *Journal of Research on Technology in Education, 37*(2), 177-195.

Khan, M. A., & Gorard, S. (2012). A randomised controlled trial of the use of a piece of commercial software for the acquisition of reading skills. *Educational Review, 64*(1), 21-35.

Kim, A., Vaughn, S., Klingner, J. K., Woodruff, A. L., Reutebuch, C. K., & Kouzekanani, K. (2006). Improving the reading comprehension of middle school students with disabilities through computer-assisted collaborative strategic reading. *Remedial & Special Education, 27*(4), 235-249.

Kreskey, D. D. (2012). Headsprout early reading for students at risk for reading failure. (Ph.D., Georgia State University). *ProQuest Dissertations and Theses.*

Longberg, P. O. (2013). Evaluation of imagine learning english, a computer-assisted instruction of language and literacy for kindergarten students. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 74.*

Lonigan, C. J., Driscoll, K., Phillips, B. M., Cantor, B. G., Anthony, J. L., & Goldstein, H. (2003). A computer-assisted instruction phonological sensitivity program for preschool children at-risk for reading problems. *Journal of Early Intervention, 25*(4), 248-262.

Macaruso, P., Hook, P. E., & McCabe, R. (2006). The efficacy of computer-based supplementary phonics programs for advancing reading skills in at-risk elementary students. *Journal of Research in Reading, 29*(2), 162-172.

Macaruso, P., & Rodman, A. (2009). Benefits of computer-assisted instruction for struggling readers in middle school. *European Journal of Special Needs Education, 24*(1), 103-113.

Macaruso, P., & Rodman, A. (2011a). Efficacy of computer-assisted instruction for the development of early literacy skills in young children. *Reading Psychology, 32*(2), 172-196.

Macaruso, P., & Rodman, A. (2011b). Benefits of computer-assisted instruction to support reading acquisition in english language learners. *Bilingual Research Journal, 34*(3), 301-315.

Macaruso, P., & Walker, A. (2008). The efficacy of computer-assisted instruction for advancing literacy skills in kindergarten children. *Reading Psychology, 29*(3), 266-287.

Mathes, P. G., Torgesen, J. K., & Allor, J. H. (2001). The effects of peer-assisted literacy strategies for first-grade readers with and without additional computer-assisted instruction in phonological awareness. *American Educational Research Journal, 38*(2), 371-410.

Mitchell, M. J., & Fox, B. J. (2001). The effects of computer software for developing phonological awareness in low-progress readers. *Reading Research and Instruction, 40*(4), 315-332.

Mostow, J., Aist, G., Burkhead, P., Corbett, A., Cuneo, A., Eitelman, S., et al. (2003). Evaluation of an automated reading tutor that listens: Comparison to human tutoring and classroom instruction. *Journal of Educational Computing Research, 29*(1), 61-117.

Nicolson, R. I., Fawcett, A. J., & Nicolson, M. K. (2000). Evaluation of a computer-based reading intervention in infant and junior schools. *Journal of Research in Reading, 23*(2), 194.

Paterson, W. A., Henry, J. J., O'quin, K., Ceprano, M. A., & Blue, E. V. (2003). Investigating the effectiveness of an integrated learning system on early emergent readers *Reading Research Quarterly, 38*(2), 172-207.

Pokorni, J. L., Worthington, C. K., & Jamison, P. J. (2004). Phonological awareness intervention: Comparison of fast ForWord, Earobics, and LiPS. *The Journal of Educational Research, 97*(3), 147-157.

Powers, S. & Price-Johnson, C. (2006). *Evaluation of the Waterford Early Reading Program in kindergarten 2005-06*. Creative Research Associates. Tuscon, AZ.

Rasinski, T., Samuels, S. J., Hiebert, E., Petscher, Y., & Feller, K. (2011). The relationship between a silent reading fluency instructional protocol on students' reading comprehension and achievement in an urban school setting. *Reading Psychology,32*(1), 75-97.

Rehmann, R. (2005). The effect of Earobics (TM) step 1, software on student acquisition of phonological awareness skills. (D.Ed., University of Oregon). *ProQuest Dissertations and Theses.*

Reutzel, D. R., Petscher, Y., & Spichtig, A. N. (Oct 2012). Exploring the value added of a guided, silent reading intervention: Effects on struggling third-grade readers' achievement. *The Journal of Educational Research,105*(6), 404-415.

Rouse, C.E., & Krueger, A.B. (2004). Putting computerized instruction to the test: A randomized evaluation of a 'scientifically based' reading program. *Economics of Education Review*, 23, 323–338.

Rupprecht, P. S. (2003). A study of the effects of a neurocognitive-based reading instruction program. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 64*(10-A), 3634.

ShelleyTremblay, J., & Eyer, J. (2009). Effect of the reading plus program on reading skills in second graders. *Journal of Behavioral Optometry, 20*(3), 59-66.

Stevenson, H.J. (2006, November). *Evaluation of the Waterford Early Reading program in pre-kindergarten 2005-2006*. San Francisco, CA: University of the Pacific.

Tracey, D., & Young, J. (2007). Technology and early literacy: The impact of an integrated learning system on high- risk kindergartners' achievement. *Reading Psychology*, *28*(5), 443-467. doi:10.1080/02702710701568488

Troia, G. A. (Nov-Dec 2004). Migrant students with limited english proficiency: Can fast ForWord Language ™ make a difference in their language skills and academic achievement? *Remedial and Special Education*, *25*(6), 353-366.

Troia, G. A., & Whitney, S. D. (2003). A close look at the efficacy of fast ForWord language for children with academic weaknesses. *Contemporary Educational Psychology, 28*(4), 465

Valliath, S. C. (2002). An evaluation of a computer-based phonological awareness training program: Effects on phonological awareness, reading and spelling. (Ph.D., Northwestern University). *ProQuest Dissertations and Theses*

Wijekumar, K. (., Meyer, B. J. F., & Lei, P. (2013). High-fidelity implementation of web-based intelligent tutoring system improves fourth and fifth graders content area reading comprehension. *Computers & Education*, *68*, 366-379. doi:10.1016/j.compedu.2013.05.021

Wijekumar, K. K., Meyer, B. J. F., & Lei, P. (2012). Large- scale randomized controlled trial with 4th graders using intelligent tutoring of the structure strategy to improve nonfiction reading comprehension. *Educational Technology Research and Development*, (6), 987-1013. doi:10.1007/s11423-012-9263-4

Wild, M. (2009). Using computer-aided instruction to support the systematic practice of phonological skills in beginning readers. *Journal of Research in Reading, 32*(4), 413-432. doi:10.1111/j.1467-9817.2009.01405.x

Woodward, A. W. (2005). The effects of breakthrough to literacy on the phonological awareness skills of students in early elementary school. *Dissertation Abstracts International Section A: Humanities and Social Sciences,68*(5-A), 1825.