A RESEARCH SYNTHESIS OF ABILITY GROUPING FOR ELEMENTARY READING INSTRUCTION

A Dissertation

Presented to

The Faculty of the Curry School of Education

University of Virginia

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

by

Susan F. Thacker-Gwaltney

B.A. College of William & Mary, 1988

M.Ed. College of William and Mary, 1989

December 18, 2014

McGuffey Reading Center

Curriculum & Instruction and Special Education

Curry School of Education

University of Virginia Charlottesville, Virginia

APPROVAL OF THE DISSERTATION

This dissertation, *A Research Synthesis of Ability Grouping for Elementary Reading Instruction*, has been approved by the Graduate Faculty of the Curry School of Education in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Mary P. Abouzeid, Advisor

_____

Marcia Invernizzi

_____

John Lloyd

_____

Catherine Brighton

_____ Date

ABSTRACT

The use of ability-based homogeneous groups for reading instruction in the elementary classroom is a widespread practice in U.S. schools (Chorzempa & Graham, 2006). This study describes a research synthesis (Shanahan, 2001) examining the use of ability grouping for reading instruction in K-5 classrooms from 1987-present. Previous reviews (Barr & Dreeben, 1996; Elbaum, Vaughn, Hughes & Moody, 1999; Kulik and Kulik 1987,1992; Lou et al.1996; Lou, Abrami & Spence, 2000; Slavin, 1987) are over 15 years old and focused on within-class ability grouping. Despite the robust special education literature advocating the importance of small-group intervention (Elbaum et al., 1999), small-group instruction in the general classroom has received less scrutiny (Barr & Dreeben, 1996). The current study is a research synthesis that examines studies of comparisons between within-class and between-class ability grouping, i.e., flexible cross-grade grouping within the general education K-5 classroom. I analyzed current support for popular organizational formats used during small-group instruction, e.g., guided reading, Daily Five, PALS electronic plans, differentiated instruction, etc. The findings provide guidance for future research on small-group reading instruction and for elementary administrators and teachers who employ small-group reading instruction.

Keywords: research synthesis, ability grouping, elementary reading instruction, within-class grouping, between-class grouping

DEDICATION

To Ben, Sophie, and Amos

bedrock, wisdom, and curiosity,

my three spinners

To my mother and father

Who gave me my first stories,

I am one of you forever (Fred Chappell)

ACKNOWLEDGEMENTS

This dissertation is the final step of a long, long journey and would not be possible with the unflagging support of all the people who have stood with me to encourage me to finish. First, a special thanks goes to Mary Abouzeid, the chair of my committee. I have driven you to wit's end, but always, you stand in my corner. I appreciate your unflagging support, steadfast enthusiasm, mother-henning, and ready pen. Your passion and commitment to literacy education and outreach to Virginia's teachers continue to amaze and inspire me. You and Marcia Invernizzi were my first mentors when I arrived at UVA, and I have learned what it means to be a reading specialist and a researcher from you. Thank you for allowing me to survey the reading field perched on your broad shoulders.

Second, a heartfelt thanks goes to each of my committee members, Marcia Invernizzi, John Lloyd, and Catherine Brighton. Thanks to Marcia Invernizzi for helping me settle on my topic and putting me on the writing path. I admire your keen intellect and passion for reading and appreciate your support of all my efforts. Thanks to John Lloyd for his unwavering confidence in me, helpful detailed edits, suggestions, and insightful questions that challenged me to stretch and grow. Through seminars and courses, John taught me what it means to be a scholar and teacher. Thanks to Catherine Brighton for being willing to help me understand the nuances of differentiated instruction, sharing resources to shape my conceptual framework and research questions, and your willingness to share your time and expertise. Each committee member was generous with her/his time and expertise, and for that I will always be grateful.

I appreciated so much the help and friendship of my volunteer coders: Beth Estill, Lorie Kolbicka, and Teresa Vedder. I could not have completed this study without their content expertise, cheerful commitment, and steadfast encouragement. You continue to show me

everyday what it means to be a reading specialist and how to work with children and teachers effectively.

Thanks to my supportive friends and talented colleagues who gave me encouragement, advice, and a kick in the pants when needed: Ottilie Austin, Julie Gray, and Carrie Simkin. I value your collective wisdom and friendship. The teachers of Virginia and beyond are lucky to benefit from your knowledge and experiences.

Finally, I would never have completed this dissertation without the love and support of my family. Thank you Ben, Sophie, and Amos for your resolute support and insistence that I finish the final hurdle. I could not have completed the last stretch without your love, patience, efforts and sacrifices. My parents, Carolyn and Jimmy, and my in-laws, Doris and Atwill, have been my staunchest supporters and fan club. Thank you for standing with me and reminding me of why I needed to finish.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Reading is a foundational skill for later academic success. Learning to read accurately, smoothly, and with understanding is an important milestone marked during the elementary grades. An emphasis on teacher-directed reading instruction is a standard part of elementary curricula in the U.S., and an accepted practice is to organize students into flexible small-groups by instructional reading level or other reading-related skills. Walk into a primary grade classroom during language arts instruction, and it is common practice to see classrooms where small-groups of children are seated at a table with an adult, reading texts or completing phonics or written activities together. Observational studies of effective schools suggest that effective classrooms maximize small-group instructional time and utilize ability-based groups (Juel & Minden-Cupp, 2000; Taylor, Pearson, Clark & Walpole, 2000).

Ability grouping refers to the practice of organizing classrooms into smaller groupings based on aptitude or achievement or ability. The use of ability groups will "directly influence the design of instruction, which in turn affects achievement" (Barr & Dreeben, 1983, p. 74). Historically, ability groups have been used for classroom assignments, across grade levels and within a single classroom.

During the last century in the United States, the use of ability groups for reading instruction in the elementary grades has been widespread but not without controversy. Classroom teachers have reported using ability grouping for instruction as a common

instructional practice since the late 19th century (National Education Association, 1968; Otto, 1950; Wilson & Schmits, 1978).

Between-class ability grouping is the practice of assigning group membership across grade levels. The terms *between-class, inter-class,* and *cross-grade* may be used interchangeably to refer to the practice of grouping students across grade levels. According to Newport (1967), the first published description of between-class grouping for reading was known as "circling" (Russell, 1946). Comparing between-class groups with "intra-class" groups, Russell reported no significant differences between the achievement in the groups.

Far less common than within-class grouping, the earliest example of inter-class grouping that rose to national attention was in Joplin, Missouri in the 1950's (Floyd, 1954). Known as the Joplin plan, named after Cecil Floyd, an assistant superintendent of schools in Joplin, Missouri, this cross-grade plan targeted students in grades 4-6. Students were regrouped across grade levels for reading instruction only and assigned to multi-grade classrooms daily for one hour. Curriculum materials were varied based on the needs of the low, middle, and high achieving students, and the plan also included a 1:1 tutoring component. Although not the first time between-class grouping had been used, the Joplin plan was the most successful and well-known at the time (Newport, 1967).

Studies of the Joplin Plan produced mixed results in terms of reading achievement. The majority of studies using comparison groups demonstrated steady gains in reading but found no significant differences from students taught using traditional within-class groups, finding only that teachers' attitudes towards the Joplin plan were more favorable (Carson & Thompson, 1964; Powell, 1964; Ramsey, 1962). A few studies reported that the Joplin plan outperformed within-class groupings (Green & Riley, 1963; Morgan & Stucker, 1960). Between-class groupings have rarely been used or studied and reported in the literature since the 1960's. The notable exception

is the Success For All program (Slavin, Karweit, Livermon, Madden, & Dolan, 1990) that will be discussed in more detail in Chapter 2.

Critics of classroom ability-based reading groups have argued against the use of ability groups on the grounds that they promote inflexible group membership, adhere to a rigid sequence of core texts, offer few opportunities for higher-level thinking skills and advancement to members of less advanced groups, and lower students' self-esteem and engagement for reading (Abadzi, 1985; Calfee & Brown, 1979; Hiebert, 1983; Slavin, 1995).

Statement of the Problem

National tests of reading achievement suggest that while reading scores have improved over the past twenty years, many students are failing to acquire the basic reading skills needed to be successful in elementary school (U.S. Department of Education, 2013). The National Assessment of Educational Progress (NAEP) reported in 2013 that 68% of fourth grade students scored at or above Basic in their reading performance. Virginia surpassed the national scores with 74% of fourth grade students scoring at or above Basic. Although Virginia's results ranked among the ten highest-scoring states in the nation, these results suggest that even in a high performing state, 26% of 4th graders lacked basic grade level reading expectations (U.S. Department of Education, n.d.).

Without the ability to read well, students face a tentative path through the middle grades and high school. Hernandez (2012) analyzed data from the National Longitudinal Survey of Youth (NLSY79) spanning 1979-1989 and reported that 16% of children who did not read proficiently by the end of third grade did not graduate from high school on time. This rate was four times higher than for proficient readers. Add concentrated poverty to the mix and 35% of poor children who were not reading on grade level by third grade failed to graduate from high school.

Today's best-selling core reading programs promote the use of small-group instruction in the general classroom. For example, major publishers, e.g., Houghton Mifflin/Harcourt Journeys, McGraw Hill, Pearson/Scott Foresman Reading Street, market leveled readers and guides for implementing small-group instruction. Although all core programs claim to be based upon research-based principles, most lack conclusive empirical support that the programs will impact student reading achievement (Allington, 2011; What Works Clearinghouse, 2007). In a review of core reading and supplemental programs, Slavin (2010) found seven quasi-experimental studies, and of those studies, none reported effect sizes above +.20 for the use of five well-known commercial reading programs.

The use of homogeneous ability-based groups for classroom reading instruction has been a common practice both historically and in an increasing number of today's elementary classrooms (Austin & Morrison, 1963; Goodlad, 1984; Loveless, 2013); however, policy and public opinion on the use of ability-based reading groups in the elementary grades has been a contentious issue. Opponents of ability grouping (Lleras & Rangel, 2009; Oakes, 1992) have confused the use of ability-based instructional groups with *tracking*, a practice that promotes academic and social inequity (Oakes, 1985). Historically, students assigned to the lowest tracks suffered due to lower expectations, less-challenging instruction and fewer opportunities for improvement and advancement. Public outcry fueled policy changes that forced many educators to rethink how they organized students for instruction. Although ability-based classrooms are rare at the elementary level, forms of tracking continue in most American middle and high schools. This research synthesis focused on the use and format of ability-based groups used within and between-classes for elementary reading instruction.

Significance of the Problem

Although there have been many studies examining the characteristics of effective schools, classrooms, and teachers for primary reading instruction, the organization, format, and quality of small-group reading instruction within the general education classroom has received less scrutiny (Ankrum & Bean, 2007). According to Barr and Dreeben (1991), "Although it has long been the wisdom in the field of reading that children learn best in small homogeneous groups, little systematic evidence exists to support or refute this practice." (p. 899)

Where general education lacks specific guidance for homogeneous small-groups used during classroom reading instruction, there is a robust literature of intervention studies for supporting struggling readers using small-group formats (Blachman, Tangel, Ball, Black, & McGraw, 1999; Bradley & Bryant, 1983; Foorman et al., 1997; Scanlon & Anderson, 2010; Taylor, Hanson, Justice-Swanson, & Watts 1997; Taylor, Short, Shearer & Frye, 1995). Many reading interventions describe targeted instruction with homogeneous groups of students. Reviews of intervention literature suggest that small-groups may be as effective as 1:1 individualized instruction for students with disabilities (Elbaum, Vaughn, Hughes, & Moody, 1999; Polloway, Cronin, & Patton, 1986). Despite a rich body of support for small-group intervention, however, some researchers suggest that small-group differentiated instruction may rarely occur in special and general education settings (Moody, Vaughn, & Schumm, 1997). The proliferation of professional teacher resources that target small-group instruction suggests that classroom teachers use a wide variety of approaches.

A renewed interest in the use of small-group instruction features prominently in Response to Intervention (RtI) frameworks. The 2004 reauthorization of the Individuals with Disabilities Education Improvement Act (IDEIA) provided the legal authority for local educational agencies to select RtI as one option for identifying and supporting students with learning disabilities. RtI

is a conceptual framework for organizing instruction and intervention in schools. It promotes early identification and remediation of learning difficulties through a multi-tiered framework of instruction that offers ever-increasing supports and intervention for students with learning and behavioral issues. Small-group ability-based instruction is a cornerstone of the intervention tiers (Gersten et al., 2008).

To be successful, RtI rests on the premise of high-quality instruction in the general classroom; the federal mandate provides no guidance on what constitutes effective general classroom instruction. IDEIA offered the states and LEAs the option of implementing an RtI framework in place of or along with discrepancy formulas to identify students with learning disabilities. The number of districts implementing the RtI strategy has rapidly expanded since IDEIA's passage. In 2010, 61.2% of school districts reported using some form of RtI within their schools with the majority using RtI primarily for reading instruction (Spectrum K12 School Solutions, 2010). As the use of RtI continues to expand, educators should expect to see the expansion of small-group instruction in the elementary grades and beyond.

The prevalence and use of small-group instruction for reading in the elementary grades has varied considerably since the 1980's. Coldiron, Braddock, and McPartland (1987) reported the widespread use of ability-based groups for reading in a study of Pennsylvania students, where over 90% were placed in reading groups in the primary grades during 1985-1986. During the 1990's, a backlash against the use of ability grouping for reading was evident in a national survey of 1,027 kindergarten teachers in which researchers reported 52% of teachers indicated that whole-group instruction was their primary mode for language arts instruction (Baumann, Hoffman, Duffy-Hester, & Moon Ro, 2000). Since 2000, there are indicators that the use of homogeneous small-groups for reading instruction is on the rise again (Loveless, 2013). In a national survey of 222 kindergarten through third-grade teachers, Chorzempa and Graham

(2006) reported that 58% of those surveyed reported using within-class ability-based groups for reading. Along with the increase in the use of ability-based small-groups in general education, teachers self-reported widespread variation in the format and practices used during small-group instruction(Ford & Opitz, 2005).

Along with RtI, the adoption of national standards for learning, the Common Core State Standards (CCSS), has drawn attention to how teachers organize and plan small-group instruction for reading. The National Governors Association Center for Best Practices and the Council of Chief State School Officers led the development of this set of K-12 learning goals. As of September 2014, the CCSS have been adopted by 43 states and the District of Columbia. Virginia is one of six states that did not adopt the standards (National Governors Association Center for Best Practices, 2010).

The CCSS have been controversial, and in particular, the issue of text selection and readability across grades 2-12 may influence how teachers organize and plan ability-based small group instruction. Proponents of the CCSS have pushed for the use of challenging texts during reading and English instruction during the primary grades and a renewed focus on teaching students texts versus teaching them reading skills (Coleman & Pimentel, 2012; Shanahan & Duffett, 2013). Opponents and researchers have argued that using texts traditionally considered at a student's frustration level to teach reading is not a sound instructional practice (Hiebert & Mesmer, 2013; Pearson, 2013). Evidence gleaned from an integrative review on ability grouping could inform these debates.

Since the late 1980's, there have been seven published comprehensive reviews on ability grouping in reading (Barr & Dreeben, 1991; Elbaum et al., 1999; Kulik & Kulik 1987,1992; Lou et al, 1996; Lou, Abrami & d'Apollonia, 2001; Slavin, 1987). The most current review was published almost fifteen years ago. All of them analyzed studies done prior to the

implementation of RtI and CCSS. The review by Elbaum et al. (1999) focused on instruction in special education. In addition, there have been no published comprehensive reviews examining the organizational structures used during ability grouping in the general elementary classroom, i.e., the "how" and "what" teachers do. This research synthesis describes the current body of knowledge on the uses of ability-based small-group instruction for reading in order to summarize the current empirical support for within- vs. between-class formats. To extend the work of previous reviewers, this study compares the findings on widely-used organizational formats for homogeneous small-group reading instruction.

Conceptual Framework

To understand why a teacher would opt to group students by ability during small-group reading instruction, it is important to know how children learn to read. Throughout the twentieth century, researchers have proposed stage models of reading based on the premise that children must pass through a successive series of steps or stages in order to learn the English writing system.

Chall (1983) proposed six stages that described how individuals progress from non-readers to mature reading adults. The stages grew out of her seminal research project on beginning reading approaches discussed later in this chapter. Chall's contributions about the important role of phonics in learning to read influenced researchers like Gough and Tunmer, who proposed a model of how children decode and understand words.

Gough and his colleagues theorized that the "simple view of reading" explained learning to read as a product of decoding and linguistic comprehension (Gough & Hillinger 1980; Gough & Juel, 1991; Gough & Tunmer, 1986; Hoover & Gough 1990). Essentially, this theory put forth a two-stage model where children build a visually-based lexicon first before passing into a decoding-based or "cipher" reading phase.

Ehri (1980, 1991; Ehri & Wilce, 1985) proposed a different stage theory of reading that focused attention on how children learn to read and write words. Spear-Swerling and Sternberg (1996) proposed a model with five stages as a "road-map" to proficient reading. They focused on pitfalls at each stage where struggling readers might require extra support to keep from going "off-track" and being identified for special education services.

Many stage models of reading share common features and terminology, but there is not a single commonly accepted model. This study rested on the stage theory of reading that originated at the University of Virginia's McGuffey Reading Center based on student observations in the school's reading clinic directed by Edmund Henderson from 1969-1989. Inspired by Charles Read's and Carol Chomsky's work with young children's writing, i.e., "invented spelling", Henderson led his students and colleagues to research and develop a model of developmental word and spelling stage theory. Eventually, Virginia's original three-stage model that described characteristics and instruction of readers was expanded to include five stages: Emergent, Beginning, Transitional, Intermediate, and Advanced readers (Bear & Barone, 1998; Bear, Invernizzi, Templeton, & Johnston, 2012).

Drawing from psychologists' theories of cognitive development and learning, the UVA reading stage model rests on the assumption that all children make optimal progress in reading if taught near their zone of proximal development (Vygotsky, 1987) with carefully scaffolded instruction (Wood, Bruner, & Ross, 1976). At the heart of this model is pairing children with instructional level texts. Children learn how to read with carefully planned developmental word study and writing. Children's instructional reading and word study (spelling) levels are assessed based on individual and combined assessments of children's language, reading, spelling, and writing skills.

Clinicians have trained generations of reading specialists and classroom teachers to apply the Virginia stages of reading to struggling readers and trained them in research-based techniques for intervention and instruction. Although primarily modeled and taught to teachers through supervised 1:1 tutorials, those lessons teach educators how to plan and apply the Virginia stage model of reading to small-group instruction. Group assignment is based on children's assessed reading, spelling, and writing needs, i.e., ability-based small-group reading instruction.

Central to the UVA stage model is a tenet that individual components of literacy, i.e., reading, developmental spelling, and writing, develop in synchrony. Due to synchrony, one component can be used to leverage information about another component in the reading stage. For example, a child's spelling can be used to predict concept of word attainment, or a child's instructional reading level can be compared to his/her spelling stage (Bear, 1991). See Table 1 for a summary of the model.

Table 1

Virginia Stage Model of Reading

Stage	Reading	Writing	Spelling
Emergent Readers, Writers, Spellers (approximately pre-K to Kindergarten)	Lacks voice to print match, i.e., concept of word May not attend to print in books & relies heavily on pictures Uses pretend reading or memory reading May recognize name or some environmental print (e.g., Chick-fil-a logo)	Is an Emergent speller Spellings bear no sound-symbol relationships (e.g., scribbles, symbol salad, random letter/numeral strings) Incomplete alphabet knowledge May represent salient sounds or a few letter-sounds by stage end	Attempts at writing include pretend writing/drawing (e.g., scribbles, drawings, random letters/#s, etc.) Lacks spaces between words Lacks vowels (medial sounds) Few or no words spelled conventionally Unable to read back one's own writing
Beginning Readers, Writers, Spellers (approximate gr. K-2)	Accurately points to words while reading, i.e. has firm concept of word in text Begins to develop a sight word vocabulary Uses developing letter-sound knowledge to decode words Slow, labored reading Whispers during silent reading time	Is a Letter-Name speller Consistently uses initial/final sounds Confuses medial short vowels, blends and digraphs, nasals Long vowels are mostly absent from child's writing Blends & segments words into onset-rime and/or individual sounds Learns sight words. Uses invented spelling	Attempts to match speech to print Segments or vocalizes words into sounds while writing Slow, labored effort to write sentences May need reminders about forming letters and spacing Learns to spell some high-frequency words Writes simple stories with a beginning, middle and end by stage's end.

Transitional Readers, Writers, Spellers (approximately mid-late 1st-3rd)	Begins to read more fluently & expressively May apply analogy-based or other strategies to read unknown words. Phases out pointing to words & shifts to silent reading Develops fluency and comprehension in reading	Is a Within Word Pattern speller. Has a core of known sight words. Most words with medial short vowels, blends, digraphs, and nasals are correct. Learns vowel patterns in single syllable words with accuracy and speed.	Begins writing in phrases with greater ease & fluency Uses the writing process with support Can spell many high frequency words Learns to write well-formed paragraphs & simple reports Begins to write for meaning Includes dialogue in stories
Intermediate Readers, Writers, Spellers (approx. gr. 4-8)	Focuses on comprehension, vocabulary and writing Reads fluently w/expression Reads mostly silently across many genres Independently applies strategies for repairing meaning. Begin to think critically and analyze their reading.	Is a Syllables/Affixes speller Reviews vowel patterns in multi-syllable words. Masters inflected endings and simple prefixes & suffixes Learns structural analysis of multisyllabic words	Organizes and write coherently about a topic Length & quality of writing steadily increases Uses the revision & editing process w/minimal support Learns to write various genres Develops a personal voice & style Masters basic research skills
Advanced Readers, Writers, Spellers (approximately gr. 8 and beyond)	Continues to focus on comprehension, vocabulary and writing Develops an understanding of multiple viewpoints. Acquires study skills Alters reading to fit the style/purpose of the text. Manages competing theories, facts, and interpretations in reading Learns to synthesize complex material from multiple sources.	Is a Derivational Relations speller Learns related sets of words based on spelling and meaning changes, e.g., consonant and vowel alternations Morphemic analysis of multisyllabic words Advanced study of Greek & Latin Roots Explore etymology of words	Writing approaches fluency Length & quality of composition steadily increases Uses the revision & editing process w/minimal support Writes across various genres Argues multiple viewpoints convincingly Employs sophisticated language that reflects abstract and complex ideas

Virginia's stage model of reading is comparable with other published theories of reading. Table 2 illustrates how the Virginia stages overlap and contrast with other well-known stage theories of reading.

Table 2

Virginia Stages and Other Stage Models of Reading

UVA Reading Stages	**Emergent**	**Beginning**	**Transitional**	**Intermediate**	**Advanced**
Chall (1983)	Prereading	Initial Reading & Decoding	Confirmation & Fluency	Reading to Learn (Single Viewpoint)	Multiple Viewpoints
Ehri (1997)	Prealphabetic	Partial-to-Full Alphabetic	Consolidated Alphabetic		
Gough & Juel (1991)	Stage 1: Visually-based lexicon	Stage 2: Decoding or cipher phase			
Spear-Swerling & Sternberg (1996)	Visual Cue Word Recognition	Phonetic Cue Word Recognition	Controlled Word Recognition	Automatic Word Recognition	Strategic Reading

Flesch (1955) ignited a firestorm among educators and parents with the publication of *Why Johnny Can't Read*, proposing that American schools were failing to teach students because of a reliance on whole-word or "look-say" methods instead of relying on phonics instruction. His ideas were embraced by a citizenry in the midst of a "Cold War", anxious that American children could fall behind Russian children.

A decade later, Chall's landmark research (1967) on the best methods for learning to read was published. Chall categorized early approaches to learning to read along a continuum between code-based (phonics) versus meaning-based (look-say) approaches. For programs that fell in-between, she labeled as "intrinsic phonics." She reported findings from 30 studies of early reading approaches and found an advantage for programs that used phonics over whole-word methods. Whole language theorists challenged Chall's findings and launched a series of studies to examine meaning-based approaches to learning words.

During the 1970's and 80's, Kenneth Goodman and Frank Smith championed the contextual and meaning-based approaches claiming that learning to read and write are "natural" processes akin to learning to speak and listen. This perspective to learning to read came to be called Whole Language. Proponents of whole language instruction eschewed code-based approaches and phonics in favor of literature immersion and authentic reading and writing experiences. These prolonged and embittered battles over meaning versus code-based approaches to reading came to be known as the "Reading Wars."

Alongside the Reading Wars, the United States experienced social unrest and upheaval due to the movement to end school segregation and the civil rights protests. Throughout the 1980's, public outcry erupted over the use of rigid ability-based classroom assignments, and the term "tracking" became synonymous with a form of social and racial inequity in American schools.

Oakes (1985) wrote the groundbreaking book, *Keeping Track: How Schools Structure Inequality*, decrying the negative effects of organizing students into ability-based classrooms and tracking students into rigid frameworks based on ability or IQ measures. She argued that minority and low socioeconomic status students were over-represented in the lowest tracks with few opportunities for movement, thus perpetuating persistent racial and class inequalities.

Around this timeframe, Slavin (1987) published one of the first comprehensive reviews of ability-based instructional groupings that concluded zero positive achievement gains from the use of ability-based classrooms. Oakes' and Slavin's work, along with other researchers (Gamoran,1986; Hiebert,1987) spurred public opinion and policy against the use of ability-based groupings in the classroom. Consequently, during this period, educators shift towards using whole-class or mixed ability groups for reading instruction.

The latter part of the twentieth century continued to be characterized by controversy between bottom-up (phonics-driven) and top-down (meaning based) approaches to reading acquisition, eventually dubbed the "Reading Wars" (Pearson, 2004). It is not until 2000, when the government-appointed 14-member National Reading Panel reviewed experimental studies on how children learn to read and examined which reading methods are most effective, that the controversy was put to rest.

The major findings from the National Reading Panel (2000) identified five key areas that impact learning to read: alphabetics, fluency, comprehension, teacher education and computer technology. These points gave rise to the Reading First "pillars" of reading: phonemic awareness, phonics, fluency, vocabulary, and comprehension, and the legislation's provision for teacher education. This legislation drove reading instruction in the country from 2003-2009 and continues to today.

Throughout the 1980's to the present, government-sponsored laws or funded reports have supported the use of small-group reading instruction and intervention for America's students. Examples include the reauthorization and renaming of Title I (1981), America Reads (1997), No Child Left Behind (2001), the Individuals with Disabilities Education Improvement Act (2004), and Race to the Top (2009) among others. As part of IDEIA, NCLB, and Title I, small-group reading instruction has come under increasing scrutiny; teachers are encouraged to shift back to

the use of ability-based small groups for general instruction, supplemental Title I, and special education.

The federal government's support for policy initiatives such as Response to Intervention (RtI) and Title I have renewed public interest in the use of small-group reading instruction. Past surveys indicated a preponderance of whole-group reading instruction (Baumann et al., 2000). More recent surveys of primary grade teachers report widespread use of small-group reading instruction (Chorzempa & Graham, 2006), but does this mean that ability-based small-group reading instruction deserves the public's trust?

The use of ability-based small groups for elementary reading instruction has been controversial with educators lining up on either side of the issue (Bracey, 1986; 1993; Condron, 2008; Hiebert, 1983: Slavin, 1987; Tieso, 2003) or offering alternatives for re-packaged versions of groups (Cunningham, Hall, & Defee, 1991; Fountas & Pinnell, 1996; Flood, Ladd, Flood & Nagel, 1992). Major reviews of ability grouping were conducted well over a decade ago (Lou et al., 1996, 2000), so a new review of small groups for reading instruction is timely. A new research synthesis to compare studies of ability groups for general reading instruction is needed.

Rationale for Research Design

Research synthesis was chosen as the preferred design since many reading research studies of the past 25 years have included a vast number of quantitative and qualitative designs with varying methods, populations, treatments, and outcomes. In situations where a diverse body of knowledge exists, research synthesis provides an objective method of finding out what is known and what needs to be known (Cooper, 2010).

Past studies of ability grouping for reading have used diverse treatment conditions, e.g., basal group, meaning vs. code-based groups, etc. and comparison groups, e.g., whole class, cooperative groups, mixed ability groups, etc. There has been tremendous variability across

outcome measures used for reading, i.e., phonics, comprehension, fluency, vocabulary, global measures, etc. Not only do studies tend to differ in terms of what they measure, but they also differ in terms of how they measure it. To illustrate, suppose study A measured the success of ability groups for reading based on the number of letter names identified, and study B compared ability group's outcomes based on the number of comprehension questions correctly answered. Side-by-side comparisons of Study A and B would end as a misleading and useless exercise if we attempted to add their results together. This variability within and across studies makes it difficult for most reviewers to look at individual research reports and draw valid or meaningful conclusions.

Research synthesis solves these problems of multiples studies, treatments and outcomes by drawing tight inclusion criteria to find and evaluate comparable studies, then converting the different outcome results to a common metric, called effect sizes (ES). If we took study A and B and converted their outcomes to ES, then we *might* be able to say something about how they are alike (or different!), whether one study had more impact, and if so how much and in what direction was the change. Furthermore, if we can gather as many studies as possible together that share common themes and outcomes, then research synthesis provides an objective way to evaluate the body of knowledge to determine whether there were any meaningful effects. This is a gross oversimplification of the process and is meant to illustrate why research synthesis was the appropriate design method for this study. See Cooper (2009), Cooper et al., 2010, or Lipsey and Wilson (2001) for a detailed handling of this topic.

Research Synthesis vs. Narrative Review

A fundamental tenet of science is that the body of evidence for theories is built through the steady accumulation of replicable well-designed studies (Shanahan, 2001). In the past, a

content expert would write a traditional literature review to summarize the history and key research findings; however, the narrative review is subject to limitations.

A narrative review may be subject to bias due to subjectivity, insufficient search parameters and opaque methods for study selection which may result in a description of a few studies that over- or under-represent the findings for any given area (Borenstein, Hedges, Higgins, & Rothstein, 2009). Narrative reviews generally do not describe their search and inclusion criteria in ways that can be repeated. The quality and impact of selected studies are judged with subjective criteria that may or may not measure the studies' effects. Studies with small sample sizes may be overlooked despite making a meaningful contribution. Studies with large samples or significant effects may be over-emphasized at the expense of smaller better-designed studies. Regardless of the type of method chosen, both narrative reviews and research syntheses require a content expert with comprehensive knowledge of the field and knowledge of the research designs used to judge study quality and impact.

Common labels for research synthesis are systematic review, research review, quantitative review, integrative literature review and meta-analysis (Valentine, 2014). Some researchers distinguish meta-analysis from research synthesis by describing meta-analysis as a statistical technique for summarizing the effect sizes of reviewed studies (Whittemore, 2005, Borenstein et al., 2009), while others use the term elastically with systematic reviews and research synthesis. The next section provides a brief history and rationale for the use of meta-analysis and research synthesis in the social sciences.

Origins of Research Synthesis

Beginning with Karl Pearson's analysis of clinical studies of the mortality rates of British soldiers who volunteered to receive the typhoid vaccine, scientists have learned that analyzing sets of studies on a given topic is a systematic way to synthesize what is currently known and

unknown for a given topic (Chalmers, Hedges & Cooper, 2002; Cochran, 1937; Fisher, 1932; Simpson & Pearson, 1904). This approach gained wide acceptance in the social sciences after William Glass (1976) introduced the term, "meta-analysis," or a reliable, quantitative study of studies, in order to corral an immense, ever-growing body of knowledge.

Glass wrote that "reading is a prime area for research integration because of the standardization of outcome measures: reading speed, comprehension, and attitude constitute the basic outcomes" (Glass, 1976, p. 7). Bus, Ijzendoorn, and Mol (2004) endorsed the process, quality and standards for conducting a meta-analysis in the field of reading and suggested:

> Even if replications of meta-analysis yield the same results, they will never constitute the final argument in the spiral of scientific research. On the contrary, the most fruitful replications will lead to new hypotheses for further study (Eagly & Wood, 1994). (p. 271)

Research synthesis is useful whenever a body of empirical studies exists for a given topic with widely varying treatments and findings.

Effect Sizes

Empirical studies are preferred for a research synthesis for the calculation of effect sizes. In order to apply meta-analytic techniques, the reviewer must find studies that share or measure common underlying constructs or outcomes in order to make meaningful comparisons. By converting a study's outcomes into common quantifiable measurements called effect size, the reviewer can compare the average magnitude of effects across studies. Statistical analysis of effect sizes allows the reviewer to gauge the statistical significance, magnitude of the findings, i.e., 'are these effects meaningful?' and whether the findings are consistent across a wide body of studies (Borenstein et al., 2009). Generally, results from different research designs and outcomes are not compared, e.g., observational studies of reading attitudes versus an empirical

study of an intervention on reading achievement would not provide a meaningful comparison (Lipsey & Wilson, 2001). For the purposes of this synthesis, the term "meta-analysis" refers to its narrower definition, the statistical calculation and summary of empirical research reports. Since this study yielded both quantitative and qualitative studies, the term "research synthesis" will be used to refer to the type of research design used for this study.

A research synthesis provides a framework and process to offer new evidence and explore issues that may not have been addressed in primary studies. One limitation of a research synthesis is that it does not allow the researcher to establish direct causality links like a well-designed empirical study; however, it may contribute new evidence to the field by exploring variables of interest overlooked by primary studies, discovering unnoticed connections and relationships, and offer new questions and targeted direction for primary reading research (Cooper, 2009, p. 33-34).

Research Synthesis Design Issues

Critics of research synthesis designs may ask: (a) how can a reviewer compare studies that differ radically on a range of treatments, subjects, setting, and outcomes? (b) how does the reviewer make certain that the synthesis has found all of the relevant information and not omitted important outcomes? Categorizing widely different study designs and outcomes is akin to the process of comparing "apples with oranges" (Rosenthal, 1979).

Research synthesis design incorporates safeguards to allow comparison across studies. First, the design requires a content expert, preferably a team, with deep knowledge of the field and research design. Second, criteria for the search and selection of studies to include and exclude must be clear so that others may repeat the synthesis if desired. Third, the synthesis design relies on effect sizes to make comparisons, even between different kinds of studies. Effect sizes gauge the strength and direction of a relationship between variables of interest as

long as they do not differ on fundamental construct of interest. Researchers have spent considerable time debating the use and formulas for finding effect sizes; however, the basic idea is to compare the difference between the treatment and the control group means divided by the control group's standard deviation (Glass, McGaw, & Smith, 1981). When studies do not provide them, effect sizes can also be estimated using *t*, *F*, or *p*-values.

Another potential flaw in research synthesis design is determining how many studies are "enough" to draw conclusions about a topic. Technically, only two studies are required to compare effect sizes, but this exercise would be of little or no value. The design of a research synthesis strives to show that exhaustive search parameters have located an entire population of studies for the given topic. The methods of mining the data must also be clear and precise as a protection against data censoring and data omission. The reviewer must clearly describe the inclusion and exclusion criteria, so that nothing that might influence the results is omitted unintentionally or on purpose. The research synthesis addresses the role and use (or not) of unpublished data, or file-drawer problem, through applying inclusion criteria and study design/rigor thresholds. Studies may be unpublished because of design flaws or due to the lack of significant results. Even if they are not used in the quantitative analysis, it is important to provide sound reasons for excluding any data.

Research synthesis and traditional narrative reviews distill the best studies to summarize a topic. Unfortunately, narrative reviews may be subjective by failing to describe their search and selection procedures. They may emphasize larger studies with small effect sizes at the expense of smaller ones with stronger outcome. The vote counting methods of significant versus non-significant results used in narrative reviews are inaccurate and misleading since they do not take into account sample size (Bornstein et al., 2009). The strength of a research synthesis approach is

its ability to conduct a meticulous search, selection, and analysis process with a goal of transparency.

Since ability grouping for reading is controversial, the topic lends itself to the use of research synthesis, as these methods allowed the reviewer to interpret impartially large amounts of study information and reading achievement outcomes. This study of studies examined the evidence for ability-based small-group reading instruction during the elementary grades and its impact on student reading achievement. It also looked for evidence of the formats used for small-group reading instruction, and whether some formats are more effective for student reading achievement than others.

Research Questions

This research synthesis examines studies on ability grouping for small-group reading instruction in the elementary grades from 1987 to the present for general education settings. Specifically, the study addresses three questions:

1. Since 1987, what evidence supports or refutes the use of ability-based reading groups in the general classroom in the elementary grades?
2. What evidence supports or refutes the use of between-class ability grouping, e.g., Joplin-type plan, for classroom reading instruction in the elementary grades?
3. What evidence supports or refutes the use of various formats/variations of ability-based small-group reading instruction used widely in the general classroom? Which ones have demonstrated a positive effect upon student reading achievement, e.g., differentiated instruction (Tyner, 2009; Walpole & McKenna, 2007); Four Blocks (Cunningham et al., 1991); guided reading (Fountas & Pinnell, 1996); Instructional Strategy Approach (Scanlon et al. 2010); PALS electronic lesson plans (PALS 2011); and Reader's

Café/Daily 5 (Boushey & Moser, 2006, 2009)?

To provide a context for these questions, the next chapter will review the history of ability grouping in U.S. schools and examine common instructional formats used for small-group instruction in the general education setting.

CHAPTER 2

REVIEW OF THE LITERATURE

Chapter 2 describes the prevalence and history of ability grouping in the U.S. In addition, the chapter explores the empirical support for the use of between-class, also known as cross-grade grouping, in elementary classrooms. The chapter concludes with a description of common instructional formats for small-group reading instruction used from 1987 to the present. The literature search was conducted through electronic database searches and a hand search through journals, books, and other materials. A detailed description of the literature search is found in Chapter 3.

Prevalence of Ability Grouping

Ability grouping refers to the practice of organizing students into groups for instruction based on student performance or achievement. It is a widespread practice within elementary schools for reading instruction. Ability grouping is often confounded with tracking, but they are not synonymous terms. Tracking refers to the practice of organizing group/class membership into assigned tracks based on IQ, achievement or other educational measures. The politics and effects of tracking in middle grades and beyond has been documented and hotly contested by educators since the late 1980's (Hallinan, 1994; Loveless, 1998; Oakes, 1985; Tieso, 2003).

The National Assessment of Education Progress (NAEP) administered every four years to samples of 4th, 8th, and 12th graders nationally includes a survey of participating teachers on their instructional practices. From 2000 to the present, there has been a steady increase in

reported use of ability-based groups for reading instruction. The 2013 NAEP reported near universal acceptance and use of small-group instruction in elementary K-5 classrooms for reading. In the 2013 Brown Report on Education, Loveless (2013) reported that the use of ability-based small-groups peaked in first grade and gradually declined throughout the elementary grades. The use of homogeneous ability groups for reading and English instruction disappears by the middle grades.

History of Ability Grouping in the United States

Grouping students for reading instruction based on ability, skills, or both has a long history in the United States and Great Britain. An emphasis on reading instruction for young children has been a standard part of early U.S. schooling since the use of hornbooks, primers, the Bible, and Psalters of the early 17th century, but it was not until the last hundred years that the practice of grouping students for reading instruction began (Barr & Dreeben, 1991).

18th century. Prior to the 1800's, children who were educated brought their own texts to school and learned to read using materials from their family. Schools followed an agricultural calendar based on planting and harvest seasons. During the 19th century, with the rise of factories during the Industrial revolution, public schools began to educate increasingly larger numbers of children. In order to manage larger groups, schools began to separate children into "classes" according to their competence in reading, writing, and spelling. Often, older children were assigned to monitor and teach younger students. The structure, timing and curricula of public schools varied considerably across rural and urban settings.

19th century. By the 1850's, Horace Mann led the common school movement that supported the use of regular texts, curricula, and timing (Tyack, 1974). In 1862, William Harris established the first documented ability-based classes in St. Louis. He also favored a standardized curriculum, and his efforts in St. Louis paved the way for ability grouping to be

used as an organizational format for meeting students' needs. As schools became aware of the increasing diversity of their populations, ability-based grouping for reading instruction became commonplace (Kaestle, 1983).

The publication of William Holmes McGuffey's readers between 1830-1860 also facilitated the move towards ability-based groups by providing a set of graded readers intended for students in grades 1-6. The McGuffey Readers marked one of the most widely used texts for teaching reading that lasted for almost 100 years with over 120 million copies sold between 1860-1960, comparable only to the Bible and Webster's dictionary in popularity. The graded readers relied on phonics and repetition to teach words using texts on a variety of topics related to morality, Christian spirituality, and content related topics.

20th century. By the 1930's, the first comprehensive basal reading programs emerged (Smith, 1965). These programs encouraged teachers to use the same materials with the entire class. By the 1960's, publishers offered districts the option to purchase supplementary readers to support high- and low-skill readers as well as parallel versions of series. These supplementary materials supported the use of ability-based instructional groups in order to meet the diverse needs of the classroom. Concurrently with new offerings from publishers, students in the United States during the late 60's and 70's experienced social change as the civil rights movement altered the educational landscape with the end of segregation. Despite a series of court cases, educators continued to use classroom-based ability grouping as a means to disenfranchise minorities, low SES, and ELL students. The use of ability-based groupings effectively re-segregated Hispanic and African American students within supposedly "integrated" schools (Oakes, 1985).

Dreeben and Barr have written extensively on the topic of ability grouping in reading instruction (Barr & Dreeben, 1983, 1991; Dreeben & Barr, 1988a, 1988b). Their book, *How*

Schools Work (Barr & Dreeben, 1983), is recognized as a thoughtfully designed description of the ways schools organize students for instruction. Their 1996 chapter in volume 2 of the *Reading Research Handbook* emphasized the historical, organizational, and ideological reasons for ability grouping for reading instruction. Although Barr & Dreeben (1991) considered instructional variables, their focus remained on how groups are organized and outcomes of group organization rather than examining the structure of specific instructional formats used within reading groups.

Studies of instructional grouping during the 1980's and 1990's tended to focus on the sociological and psychological aspects of instructional grouping, e.g., socioeconomic status, political equity, self esteem (Gamoran, 1989; Gamoran, Nystrand, Berends, & LePore, 1995; Rosenholtz & Wilson, 1980) or achievement outcomes of small-group instruction(Polloway, Cronin, & Patton, 1986; Kulik & Kulik, 1992; Slavin, 1987). Researchers recognized that reading instruction may account for individual student differences and called for a shift away from the study of reading and general learning processes to the study of classroom reading instruction (Allington, 1983; Duffy, 1981).

As a consequence, during the 1990's, the use of ability-based classrooms virtually disappeared, and homogeneous grouping for reading instruction dramatically declined in response. Districts, seeking to distance themselves from the use of classroom-based ability groups, adopted mastery learning, cooperative learning, non-graded classrooms, or multi-age or multi-grade classrooms. However, proponents of gifted and talented education rejected the de-tracking movement's insistence that all forms of ability-based groups be eradicated from schools. The best known and most cited are a series of reviews and studies conducted by Kulik and Kulik (1987, 1992) examining the affects of homogeneous groupings on achievement gains for low-, medium-, and high-ability achievers. Kulik and Kulik (1987) concluded that both high

achievers and low achievers demonstrated gains and benefitted from ability-based groups for classroom instruction.

In 1987, Robert Slavin published the most-cited comprehensive review of ability grouping in elementary schools. He compared the use of between- and within-class ability-based groups in grades 1-6 in the U.S. in existing studies up to 1985. His best evidence synthesis did not include the most common alternatives to ability grouping at the time: studies on mastery learning (Block & Anderson, 1975), individualized instruction, cooperative learning in mixed ability teams (Oakes, 1985), non-graded plans, open classrooms, or team teaching. Slavin's comprehensive review was different from other reviews of its kind since it compared within-class versus between-class ability grouping studies, something which other researchers had not considered before. In addition, Slavin included unpublished studies and dissertations in his review. He reported on 47 studies published between 1932-1985 that met his inclusion criteria which included:

1. Studies using ability-grouped classrooms compared to heterogeneously grouped control classrooms.
2. Studies reporting achievement data from standardized normed achievement tests. Studies could not rely on anecdotal accounts or teacher perceptions/attitudes.
3. Studies used random assignment or equivalent matching of comparable classes, e.g., by matching students through use of an IQ measure or other reported standardized measure.
4. Classrooms in studies used ability grouping for at least 1 semester.
5. Studies met a minimum sample size of at least 3 experimental and 3 control students included in each study.

Slavin gave a careful accounting of his methods to provide a comprehensive search of the literature, criteria for including and excluding studies, analysis of the key studies included, explanation of how effect sizes were calculated, and gave a systematic reporting of the results. The main findings for this best evidence synthesis were the following:

1. Zero effect size was found for classroom-based ability grouping across 14 studies, meaning that grouping students in classrooms by ability does not increase their achievement.
2. Regrouping classes for reading and math scheduled at the same time within grades based on student performance, rather than IQ, was inconclusive. None of the seven studies included demonstrated consistent gains.
3. Fourteen studies were reviewed using Joplin or Joplin-like plans for reading instruction only. There were too few math studies to be included. In Joplin plans, teachers regrouped students across grade levels for reading only. Eleven out of the 14 studies reported students made positive gains in achievement. Four of the 11 studies analyzing achievement also used controls to examine student differences. There was no consistent pattern in these four studies. Students either demonstrated consistent gains over control groups, or no significant differences were noted at all. There was no evidence that use of the Joplin plans resulted in negative achievement outcomes for students.
4. Slavin examined eight studies of within-class ability grouping, primarily for math. These studies compared ability-based groups within grade levels to non-grouped controls. Because reading instruction using grouping within grade levels was so widespread, there were insufficient studies with controls to draw any conclusions. For

> math, Slavin found support for the use of ability-based groups within classrooms. There was no evidence to indicate the ability-based groups within classrooms favored high achievers at the expense of lower achieving students.

The Slavin (1987) study concluded with four criteria for using ability grouping effectively: (a) reduce heterogeneity in the specific skill being taught; (b) use a flexible plan to respond to mis-assignment and changes in student performance after placement; (c) the pace and level of instruction must be varied to correspond to students' needs and learning rates; (d) the number of groups should be kept small to allow for adequate instruction from the teacher in within-class ability grouping plans (Slavin, 1987, p. 322).

With 957 citations in Google Scholar, Slavin's 1987 best evidence review has received the most attention of the reviews of ability grouping by both proponents and detractors of ability grouping. Gamoran (1987) reflected that although Slavin paid careful attention to methodological issues, the study did not provide details about the substantive issues, e.g., type and quality of instruction within these ability-based groups. Without this detail, the reader is unable to decide whether the different studies of ability grouping were grouped together as "fundamentally different phenomena under a single rubric" (Gamoran, 1987, p. 341).

Hiebert (1987) commented that the studies Slavin included were outdated and provided at best a historical perspective of ability grouping in elementary schools up to the mid-20th century. Many of the studies cited by Slavin were from 1960 and earlier. She argued that grouping might not account for the differences in student outcomes without careful examination of the classroom processes and contexts for learning. For example, Hiebert used the example of the Joplin plan (Floyd, 1954), noting it had additional components beyond cross-grade grouping that may have influenced its positive outcomes, e.g., professional development, specific reading methods,

materials, and time allotted for recreational reading. She concluded that Slavin's comprehensive review had little to say to classroom instruction in the 1990's.

Another widely cited review of ability grouping was a meta-analysis conducted by Kulik and Kulik published in 1992 that examined findings for five types of programs: (a) multi-level classes (56 studies); (b) cross-grade grouping (14 studies); (c) within-class grouping (11 studies); (d) enriched classes for gifted and talented students (25 studies); and (e) accelerated classes for gifted and talented students (23 studies). In contrast to Slavin's (1987) study, they examined whether ability grouping had differential effects for low-, middle-, and high-achievers. They concluded that there were clear consistent advantages to grouping for instruction, in particular that ability grouping for instruction benefitted high achievers.

Neither Slavin (1987) nor Kulik and Kulik's (1992) reviews addressed the best conditions for using within-class ability grouping, so the next substantive analysis of small-group instruction occurred with a meta-analysis of 66 studies (Lou et al., 1996). Unlike its predecessors, Lou et al. (1996) included additional studies of within-class ability groups and coded for factors such as grouping practices, instruction, and varying curriculum. In addition, they expanded the range to include elementary through post-secondary levels. Prior studies had focused on target grade ranges. Overall, they found small but positive effects for the use of within-class ability grouping across all subject areas. In particular, they reported that homogeneous groups for reading made larger gains in achievement than heterogeneous groups ($d = +.36$). Of note, they discussed that studies utilizing teacher-made informal measures had larger effect sizes than research-made, standardized measures. They noted that these differences may be the result of experimenter bias, or teacher-made tests may reflect a more sensitive measure of student achievement. Studies that varied instructional materials among groups resulted in stronger effect sizes. Also interesting was that professional development for teachers

resulted in a slightly higher and positive effect than studies where teachers and controls received no training. The authors noted that the positive effects for small-group instruction were maximized when teaching methods and materials varied for small-groups. Group size was optimal for 3-4 member teams, but effect sizes diminished when group assignments grew larger, i.e. groups of 6-10 students.

Group membership also affected outcomes. For low achievers, heterogeneous groups promoted higher outcomes, but the use of homogeneous groups supported the achievement of medium-ability students. High-achievers showed no significant differences in either grouping format. The authors were quick to point out that low achievers may benefit from the additional support and tutoring of a mixed group, but it is not reciprocal for medium and high achievers. Lou et al. (1996) cautioned that results for the low achievers may have been influenced by teachers' low expectations or poor quality instruction in the homogeneous groups.

Until the early 1990's, most grouping studies tended to focus on intervention. Elbaum, Vaughn, Hughes, and Moody (1999) reported a meta-analysis of twenty studies published between 1975-1995 that examined the results of grouping for reading intervention on students with disabilities. Their meta-analysis focused on integrating findings from reading interventions that used either student pairing, small-groups, or specific combinations of two or more grouping formats. The overall mean weighted effect size for the three grouping formats was $d = +.43$, suggesting a small but significant positive effect of group interventions. Individually, only the mean effect sizes for pairing and small-groups were significantly different from zero. Findings suggested that a focus on general reading and comprehension yielded higher effect sizes than a focus on word recognition. The authors hypothesized that general reading and comprehension measures may yield higher effect sizes than word recognition measures regardless of the focus of instruction. The authors also noted that the small number of studies available for inclusion may

have influenced the effect sizes.

Recent comprehensive studies have utilized large nationally-representative data sets, such as the Early Childhood Longitudinal Study (ECLS) data to draw conclusions about ability grouping and reading instruction during early school experiences. Condron (2008) examined the performance of low, middle, and high performing ability-based groups within the first cohort of the first- and third-grade students and found that ability-based groups favored high-achieving groups, made no difference for the middle groups, and harmed the low achievers compared to classrooms where no grouping was used. Condron argued that the unequal achievement patterns reflect the use of ability grouping. Although analysis of the large data set revealed significant differences in achievement, the interpretation focused on student outcomes with little information provided on group assignment, group format, frequency and type of grouping changes, curriculum materials, organization and format of the small group instruction. Without information on the format and quality of the instruction provided, it is difficult to draw conclusions about the use of ability groups for beginning reading instruction from this study. Condron's study could not capture the nuances of planning small-group reading instruction.

Between-Class Ability Grouping

Placing students into permanent homogeneous classrooms based solely on ability or achievement was discredited by Slavin in his 1987 comprehensive review of ability grouping in the elementary grades. He reported an overall effect size of .00 for 17 comparisons in 14 class-assignment studies. Later, within the same review, Slavin examined the support for within-class math ability groups and flexible between-classroom ability grouping based for reading instruction, i.e., Joplin plans. For the use of small groups in math, he found a small effect size, ES = +.34. The largest effects occurred for between-class grouping studies with a net benefit of +.45 standard deviations across 11 out of 14 studies, i.e., the largest mean effect size reported in

his study. Based on his research, it is no surprise that he later chose cross-grade grouping for reading as one of the components of the Success For All program (Slavin & Madden, Karweit, Livermon, & Dolan, 1990).

The only other widely cited example of between-class ability grouping is the Success For All (SFA) program (Madden, Slavin, Karweit, Dolan & Wasik, 1993; Slavin et al, 1990). According to the Success For All Foundation© (Success For All Foundation©, n.d.), SFA is used by more than 1,300 schools in 500+ districts across 48 states, Guam, and the Virgin Islands. SFA is a comprehensive school reform program that uses between-class, cross-grade grouping as one of several programmatic components. SFA employs cross-grade reading groups for 90 minutes daily along with 1:1 tutoring and differentiated teaching across grades 1-3. SFA has additional components, e.g., ongoing professional development, cooperative learning, and specific instructional curricula and learning strategies. The program has been replicated across multiple sites with varying degrees of success (Borman & Hewes, 2002; Quint et al., 2013). Although it has demonstrated consistent gains in reading achievement, SFA is an expensive program, e.g., an average of $120,000 for year one, $55,000 for year two, and $45,000 for year three for a school of 500 students. SFA requires broad initial support from teaching staff (80% buy-in) before it can be implemented in a building (SFA Foundation, n.d.).

Between-class placements for reading or math are uncommon in the elementary grades and practically non-existent in the special education literature. Gifted education researchers are the primary advocates for the use of temporary flexible groups and differentiated or parallel curricula to meet the needs of high achievers (Tieso, 2003, 2005).

Formats for Small-Group Reading Instruction

Since the 1990's, a wide variety of organizational and instructional formats for small-group instruction have flourished in the publications of professional development literature and

been featured in state, national, and local reading associations' conferences and publications. Examples of these include guided reading (Fountas & Pinnell, 1996); Four-Block (Cunningham, Hall, & Defee, 1991); Daily Five – Reader's Café (Boushey & Moser, 2006, 2014), differentiated reading instruction (Tomlinson, 1999; Walpole & McKenna, 2007), and in the state of Virginia, the Phonological Awareness Literacy Screening (PALS) electronic lesson plans (PALS, 2011). In addition, guidelines for small-group implementation accompany the programs of most commercial core reading programs. Instructional formats for small-group instruction may fall loosely into two categories: guided reading (text-centered) versus skills-based approaches (Kosanovich, Ladinsky, Nelson, & Torgeson, 2010).

The study examined the efficacy of popular small-group formats: guided reading, Reader's Café, PALS electronic lesson plans, and differentiated reading instruction (Walpole and McKenna, 2007). Each format is described below in alphabetical order. What distinguishes them from commercial core reading programs is their lack of specific instructional materials, unscripted approaches, and organization of learning around a developmental literacy continuum. All of the named approaches provide support in varying degrees to students as they learn to read and write. All of them share a recognition that what the teacher does and how instruction is implemented has a major impact on learning beyond that of a specific program (Allington, 2002; Bond & Dykstra, 1967; Tivnan & Hemphill, 2005). These approaches to small-group instruction emphasize that teacher knowledge and professional training are important and needed to implement beginning reading instruction consistently with success (Scanlon, Gelzheiser, Vellutino, Schatscheider & Sweeney, 2008).

Daily Five – Reader's Café. First published in 2006 and 2009, the Daily Five and Reader's Café are popular literacy management frameworks authored by two classroom teachers and sisters, Gail Boushey and Joan Moser, who refer to themselves as "Two Sisters." The Daily

Five refers to a classroom organizational plan for managing student independent work so a teacher might offer conferences, 1:1, or small group instruction. In their book, *Reader's Café*, Boushey & Moser (2014) include information on how to set up and run flexible within class ability groups based on needed reading strategies rather than instructional reading level.

They describe five types of independent literacy tasks: (a) Read to Self; (b) Work on Writing; (c) Word Work; (d) Read to Someone; and (e) Listen to Reading (Boushey & Moser 2009). The goal of these tasks is to build children's stamina for reading and writing so the teacher may use that time to confer with individual students or conduct small-groups. Critics of the Daily Five have argued that the literacy scheme promotes whole-group instruction followed by independent work and reduces the amount of time allocated for small group instruction.

Boushey and Moser's structure (2009) does not support meeting with the same small-groups daily. Flexible small-groups are based on criteria other than instructional reading level; students are grouped by similar reading strategies instead. For example, students may be reading different levels of books but discussing a common strategy during small-group time. Boushey & Moser (2014) do not advocate meeting with the same small-groups daily using a structured format. Although the Daily Five may provide a convenient tool for organizing independent work, the underlying premise of Reader's Café is a repackaging of whole language instruction with its focus on comprehension and meaning. Grouping students by strategy-need creates a mini workshop approach; the authors provide minimal guidance for systematic teacher-directed word identification skills.

Differentiated instruction. Differentiated instruction refers to adjustments made to the curriculum, assignments, or classroom to meet the learning needs of individual students within the classroom (Washburne, 1953; Stradling & Saunders, 1993). Gardner's (1983) seminal book on the nature of human intelligence, *Frames of Mind*, describes his theory of multiple

intelligences and brain-based learning that served as a theoretical bridge for educational researchers to assert the importance of differentiated instruction as critical for increasing student knowledge. Differentiated instruction is "an approach to teaching that advocates active planning for student differences in classrooms" (Tomlinson, 1999, 2001, 2008) and providing academically responsive instruction to students (Tomlinson et al., 2003).

Walpole and McKenna (2007) discuss the importance of differentiated reading instruction, but they distinguish their approach from others by focusing on the use of flexible small-groups that target specific reading skills based on assessed student need. Group membership relies on targeted assessments to determine whether students have mastered phonological awareness, word recognition, fluency, comprehension, and vocabulary—all necessary steps along the reading continuum.

Tyner (2009) described a series of small group lesson formats differentiated according to where the child falls along the literacy continuum. She defines eight distinct literacy stages and recommends that teachers use assessed instructional reading levels to place students into small groups. Small-group instruction is organized around fluency, sight words, phonics, phonemic awareness, and guided reading. For each stage of reading, Tyner suggests a prescribed sequence of word study loosely based on developmental spelling principles described in *Words Their Way* (Bear, Templeton, Invernizzi, & Johnston, 2014).

Four Blocks instruction. In the early 1990's, Cunningham, Hall and Defee (1991) described a yearlong first grade experiment where they provided multilevel instruction without the use of ability grouping. They described organizing a two-hour language arts time frame into four equally divided blocks that represented the "four major approaches to literacy instruction of the past century" (p. 568). The four blocks consisted of a writing block, a phonics and sight word block (word wall and working with words), a basal reading block (guided reading), and a

literature-based block (reading real books). Although the children met in small-groups for guided reading, group membership changed daily and was not contingent on performance. By year's end, all but three students in their study had made steady improvement as measured by an informal reading inventory. High achieving first grade students were reading instructionally between 3rd-6th grade. However, only 83% of the class was on or above grade level with four of the twenty-four students (16%) reading below expected levels (preprimer – primer). Despite the high number of struggling readers, Cunningham et al. (1991) described this as a successful format and indicated that the entire grade level had decided to adopt the organizational layout following the study. The authors attributed its initial success to the skill of the classroom teacher and to providing differing approaches to teaching reading. The primary emphasis of the four block structure was multi-level and multi-approaches to primary reading instruction. This format for reading instruction was popularized during the 1990's but virtually disappeared from classrooms after a scathing Fordham Center report written by Moats (2007) characterized Four Blocks as a flawed approach masquerading as a scientifically-based approach to reading instruction. The national Four Blocks website officially closed down in 2012.

Guided reading. From the mid-1980's to the early 1990's, educators moved away from the use of small-groups in the classroom, preferring instead to provide whole-group instruction. This was primarily done to avoid the perceived pitfalls of tracking and educational inequity prevalent in ability-based classrooms. In the mid-1990's, Fountas and Pinnell (1996) introduced and popularized the use of guided reading instruction as a means to reinstate the use of ability-based groups for reading instruction back into schools. To avoid the ideological pitfalls associated with tracking, they focused on instructional reading level, text selection, and using flexible grouping patterns. Guided reading has morphed since its inception from a specific instructional technique to a general term that is synonymous to some with small-group reading

instruction.

Traditional "guided reading" as described by Fountas and Pinnell (1996) consists of careful selection of leveled readers to match the text to readers' needs as assessed through running records, an informal reading inventory, or other formative measures. Guided reading brings together a small group of students who read on the same instructional level or who require support with the same reading strategy through a scaffolded experience as the teacher leads them through a text using a series of instructional moves that happen before, during, and after the reading.

One of the primary consequences of Fountas & Pinnell's work has been a collective effort by educators and publishers to create a "leveled books" gradient in order to determine text readability to aid the text selection process. The primary emphasis in guided reading is to encourage students to read texts with understanding by selecting books with an appropriate level of challenge with coaching support from an adult. Guided reading is a term that has become synonymous with small group instruction, though few realize that it is originally derived from the work of Emmet Betts (1946) (Ford & Opitz, 2011).

Interactive Strategy Approach (ISA). Scanlon, Anderson, and Sweeney (2010) published an approach to early literacy instruction called the Interactive Strategy Approach (ISA) based on a series of intervention studies conducted in the late 1980s and early 1990s (Vellutino et al., 1996; Vellutino, Scanlon, Zhang, & Schatschneider, 2008). The central goals of the ISA approach are to focus on word identification acquisition through teaching code-based and meaning-based strategies to identify unfamiliar words in texts. Initially, ISA was documented in 1:1 and small-group formats, and its primary use was in intervention settings. The authors of the ISA studies have since expanded from their initial focus on pull-out intervention to the use of interactive strategies in the general classroom setting. In a discussion of criteria associated with

small-group reading instruction, Foorman and Torgeson (2001) suggested that principles of effective small-group reading instruction are the same whether the intended goal is general instruction or intervention, i.e., the principles of effective instruction apply across settings.

PALS electronic lesson plans. In the state of Virginia, the Phonological Awareness Screening assessment (PALS) is a state-funded reading diagnostic screening and assessment tool designed for the primary grades (Invernizzi, Meier, Swank, & Juel, 2004). The PALS assessment is available for Pre-kindergarten through third-grade students and is administered by 99% of Virginia's public school districts. A Spanish version will soon be available as well as a new version for grades 4-8 that is currently being piloted for 2015. In addition to the screening tool, the PALS office provides a series of curriculum-based measures, called RtI Quick Checks, to monitor student progress during intervention support. To support classroom instruction and intervention, the PALS office offers numerous professional resources and activities to support classroom instruction, and this includes lesson-planning support for using differentiated ability-based groups for daily small-group instruction (PALS, 2011).

Chapter 2 explored the history and types of ability grouping in the U.S. in elementary classrooms. The use of small groups within and across classrooms for reading instruction in the elementary grades has had a long and contentious history in the U.S. Along with the commercial basal reader, the last twenty years has seen the emergence of popular instructional formats for ability-based small-group instruction that do not rely on specified materials, use unscripted programs, and organize learning around a stage model of reading or around specific skills/strategies for reading instruction. In the third chapter, the method for conducting the research synthesis on ability grouping will be explained.

CHAPTER 3

METHOD

Overview

Chapter 3 provides a detailed description of the method and procedures that were followed during the research synthesis. The synthesis began by defining the independent and dependent variables within this study and using them to delineate a conceptual framework of the treatments and outcomes involved in studies of small-group reading instruction that fit the inclusion criteria. Following procedures suggested by Cooper (2010); Cooper, Hedges, and Valentine (2009); Lipsey and Wilson (2001); and Shanahan (2001), a search was made for all relevant studies that addressed the three research questions. In a research synthesis, the unit of interest is a study, so instead of working with live subjects, studies serve as "participants" instead. An exhaustive literature search yielded an initial 35,584 hits that were reduced to 462 studies by screening for titles, abstracts and key terms. Applying preliminary inclusion criteria to the titles and abstracts of the 462 studies and removing duplicates decreased the set to a total sample size of 243 studies. Inclusion and exclusion criteria were applied to the full-texts of all 243 studies, yielding a total of 29 studies that met the inclusion criteria for this synthesis. Of the 29 included studies, 23 yielded effect sizes for additional analysis. Six studies met inclusion criteria but did not yield sufficient data for effect size calculation. These remaining six were set aside for qualitative review and will be discussed in more detail in the results section of chapter 4.

Three coders volunteered to be trained and assisted the reviewer with the literature screening and selection process to protect against study selection bias, data censoring and data

entry errors. Disagreements were resolved through conversation and further review. If the coders could not resolve the agreement, and the study met all but one inclusion criteria, then it was added to the analysis. The remainder of this chapter will describe each of the steps of the research synthesis in details and explores the key variables and methods used for searching, selecting, coding, and analyzing the studies in the synthesis.

Independent and Dependent Variables

A conceptual framework was used to describe the independent and dependent variables of interest that drove this study. The conceptual definitions provide a general over-arching description of the independent and dependent variables used, and the operational definitions explain the variables by using concrete examples of how they were identified during the study search and selection process (Cooper, 2009). The synthesis had two independent variables: ability grouping and small-group reading instructional format. The synthesis included one dependent variable, i.e., reading achievement. The variables are summarized in Table 3 below.

Table 3

Key Variables and Definitions

Independent Variables		
Variable	Conceptual Definition	Operational Definition

Ability grouping for reading	For the purposes of this synthesis, ability grouping refers to the practice of organizing a heterogeneous classroom into homogeneous small-groups by an educator for classroom reading instruction. Ability grouping may occur within a classroom, within a grade level, or across classrooms/grades based on student achievement or ability. Ability groups for reading instruction are dynamic and flexible in terms of frequency and duration and should not be confused with the use of ability to create a permanent class placement.	**Within-class** ability groupings assign students within a single classroom to homogeneous groups ranging from two to ten students based on instructional reading level, assessed reading strategies/skills, or other ability-based criteria. **Between-class** groupings assign students within a single grade level or across multiple grade levels (cross-grade) to homogeneous groups based on instructional reading level, assessed reading strategies/skills, or other ability-based criteria. Within and between class ability groups are for reading instruction only for the purposes of this study.
Small-group reading instructional format	The study describes specific instructional methods/activities used during within-class or between-class small-group reading instruction. It addresses the "how" and "what" the general education teacher does during any regularly scheduled teacher-directed small groups. It does not include reading-related skills taught during content area instruction, e.g., math, science, social studies or process writing.	The study may report a specific format, e.g., guided reading, Reader's Cafe, PALS electronic lesson plan, etc. or provide details of the method/materials used by the general education teacher to provide on-going reading instruction. It may describe a general orientation or routine for the use of regularly scheduled, teacher-directed small groups, e.g., differentiated instruction. Instructional format may be oriented to teaching specific reading skills or text-centered, i.e., use of leveled texts. It did not include studies that focus on a single aspect of literacy instruction for enrichment or targeted intervention, e.g., comprehension intervention, fluency instruction, or vocabulary enrichment.

Dependent Variable

Variable	Conceptual Definition	Operational Definition

Reading achievement	Refers to the assessments used to measure progress and growth in the literacy skills taught during the small-group instruction in the general classroom. It may include word identification, phonological awareness, phonics, fluency, comprehension or vocabulary work. Assessments may target specific literacy areas or provide a global measure of overall reading achievement. It does not include content area reading instruction, e.g., math, science, social studies or process writing outcomes.	Standardized tests of reading achievement, e.g., Woodcock Reading Mastery test Informal or formal measures of word recognition, spelling, or passage reading Informal or formal measures of alphabet, phonological or phonemic awareness Informal reading inventory passages Researcher-created reading tests In studies with multiple outcomes, formal measures of reading outcomes were used when available.

Synthesis Independent Variables

To answer the question whether the use of homogeneous small-group reading instruction is effective in the general education K-5 classroom, two independent variables, ability grouping and the format of small-group instruction within those groups, were examined. The term ability grouping may be widely applied to a range of content areas and grades out in the literature, but for the purposes of this review, its definition was confined to the use of homogeneous teacher-directed small-groups for reading instruction. Group membership assignment was based on one or more of the following categories: instructional reading level, assessed need/use of specific reading strategies or reading skills, or based on achievement/placement tests. In this study, ability groups does not include heterogeneous or mixed-ability small groups by definition, since these include students with varying levels of skills and achievement working together.

The second independent variable was the research support for using specific instructional formats during ability-based small-group reading instruction. For the purposes of this synthesis, small-group reading instructional format refers to the constellation of routine skills and activities used during teacher-directed small-group reading instruction, i.e., the "what" and "how" of

instruction that occurs within each small group. It is common practice for teachers to self-refer to their small-group instruction based on the type of instructional activities/skills/program that they are using, e.g., guided reading or basal reader groups, but the meaning of these terms can vary widely (Ford & Opitz, 2008).

There are numerous formats for small-group reading instruction in the regular education setting, and a search was made for empirical evidence to describe the most common ones found in the literature, on the website of the International Reading Association (IRA, n.d.), and formats encountered by the reviewer during extensive fieldwork experiences in public school districts in ten years across the state of Virginia. Evidence was sought to support the following formats: Differentiated instruction (Tyner, 2009; Tyner & Green, 2005; Walpole & McKenna, 2007), Daily Five/Reader's Café (Boushey & Moser, 2006, 2009), Four Blocks (Cunningham et al., 1998); Guided Reading (Fountas & Pinnell, 1996); Interactive Strategy Approach (ISA) (Vellutino & Scanlon, 2002); and the PALS Electronic Lesson Plans (PALS, 2011). It should be noted that during the study, two additional formats were added, and one was removed from the coding guide. Basal reading group, a generic term to apply to any commercial reading program designed for teaching children to read in the regular setting, and Success For All Roots and Wings, a curriculum and set of consistent activities and routines used during cross-grade small group instruction, were added to the search. The Interactive Strategy Approach (ISA) was removed as a category for small group instructional format during the coding process since these studies did not occur within the regular education setting as the reviewer had initially thought. The ISA studies focused solely on intervention rather than general class teaching, and reported using 1:1 instruction instead of small group formats.

Synthesis Dependent Variable

Studies of ability grouping have measured the impact of student attitudes, behaviors, perceptions, and motivation/engagement, but the ultimate issue of whether an educational practice works or not is how well a study's treatment impacts a student's reading achievement outcomes. For this reason, the synthesis focused exclusively on outcome measures related to student achievement. Studies that did not report or use reading achievement measures as an outcome were excluded.

Because reading develops in a stage-like progression with multiple constructs that influence and interact to produce a seamless result, there is no single accepted measure that is a "gold standard" used during studies of reading research. Many studies employ multiple outcome measures in order to tap the widest array of variables that can measure components of student reading growth. In the elementary grades, many students will likely pass through most of the reading stages, (refer to Chapter 1), and each reading stage has specific emphases for instruction that support children's progress towards the next level, e.g., phonemic awareness influences decoding, concept of word precedes sight word development, etc. Because of the stage-like progression and interactions among these skills, reading researchers seek to use measures that tap critical reading constructs, are decent predictors of student outcomes, and do not have ceiling effects. For example, alphabet and letter sounds are common assessments for K-1 students; however as children master these skills, they do not provide meaningful comparisons among older students due to their ceiling effects.

Reading outcomes included in this synthesis related directly to one or more of the five areas operationalized following the National Reading Panel report (NICHHD, 2000): phonics, phonological awareness, fluency, vocabulary, or comprehension. Outcomes measures were organized according to seven broad categories. These included the five areas above plus two

additional categories called "Word Identification/Attack" and "Global Reading." Word Identification/Attack included outcome measures of children's word list reading using graded or non-graded words in timed or untimed formats. Nonsense word lists were also included under this designation.

Global reading referred to any results yielded from overall or general reading batteries. Examples of global reading outcomes were standardized normed instruments, e.g., Stanford Achievement test, or Woodcock Reading Mastery test. If a study indicated that the subtest measured a specific construct, then the effect size was coded for that area. If the study did not specify how the subtest was used or reported a combined score, then the global reading achievement score was used. Only standardized achievement or researcher-designed instruments that provided evidence of reliability and validity checks were included. Informal classroom assessments or study-designed tools that lacked reliability or validity measures were noted but not used to calculate effect sizes in the analysis.

Literature Search

Following the process described by Cooper and Hedges (2009) and Shanahan (2001), the reviewer employed a literature search that included the following four tasks:

1. Conducted an electronic search through the following online databases: EBSCO, Google Scholar, JSTOR, PSYINFO, Web of Science, What Works Clearinghouse, and ProQuest Dissertations and Theses.
2. Conducted a hand search of six reading and education journals for related studies from the last ten years, 1994-2014. To document the search and aid any future researchers, a search log was used to keep track of the studies, search terms used, and results.

3. To avoid "file-drawer bias", unpublished papers and dissertations were retrieved from ERIC documents, Dissertations Abstracts. Priority was placed on published empirical studies of ability grouping within and between classrooms that have navigated the peer review process.
4. "Footnote-chasing" (White, 1994) was used to locate related secondary sources to insure that studies were not overlooked during the online database or hand search.

See Table 4 for a list of the databases used, key search terms and descriptors, fields and search options, and the total number of hits. Consultation with a librarian was used to refine terms and structure the initial search.

Table 4

Literature Search Terms

Search Source	Key Words	Fields	Search Options	# of Hits
APA PSYCH NET (PSYC INFO	"ability group", read*, "reading instruction", "elementary", *NOT* intervention, NOT math	Any Field Journal Title Abstract Peer-Reviewed Journals	Boolean Phrases Year: 1987 *TO* 2014 Peer-Reviewed Journals Only	96
EBSCOhost Research Database (using Academic Search Complete, Education Full Text (H.W. Wilson), Education Research Complete, ERIC, Psychology and Behavioral Sciences	ability group* AND read* instruction AND elementary NOT intervention NOT math ability group* AND read* instruction OR small group* OR small-group instruction AND elementary TI (grouping AND read*	All fields Title and abstract Subject terms	Published Date: 19850101-20110131 Expanders - Apply related words; Also search within the full text of the articles Search modes - Find all my search terms; also search	34,019

Collection, and Teacher Reference Center, SPORT Discos with full text	instruction OR AB small group* OR small-group instruction AND elementary) AND SU ability group* OR SU small-group instruction		within full text of articles Boolean Phrase	
JSTOR (Education, Social Sciences, Psychology, & Sociology databases)	ability grouping (title), reading instruction and elementary Slavin, R. (1987)	all text fields, abstract only selected Citation Report	1987-present	276 75
ProQuest Dissertations & Theses Global (Social Sciences)	ability grouping, reading instruction, elementary ab (ability group) AND ti((reading OR small group)) AND ab(elementary)) NOT (middle grade* OR grade 6) NOT intervention NOT math	All text fields and title Abstract only English only	1987-present	125 30
Web of Science	ability group* and reading and elementary Citation search: Slavin, R.E. (1987)	All text fields Citation Report	1987-present	140 215
Google Scholar	Search words: Ability group, read, reading	Find articles with All of the words With exact phrase In title of article	1987-present	6
Search #2: Initial search was expanded using individual separate searches with additional keywords/terms in combinations to capture studies addressing specific small-group instruction formats				
JSTOR	Four Block, guided reading, elementary, reading achievement Daily Five, Reader's Café, elementary, reading instruction PALS lesson plan, Phonological Awareness Literacy Screening, electronic lesson, read*, small group	Item title and abstract	1987-present	107 29 0 28

	TI success for all or Joplin plan AND AB reading achievement			
EBSCO	Daily Five, Reader's Café, elementary, reading instruction Four Block, reading instruction "guided read*", elementary, read* achievement TX phonological awareness literacy screening AND reading achievement AND AB small-group instruction NOT TI intervention TI success for all AND AB reading achievement	Title and abstract	1987-present	49 1 88 23 37
Web of Science	ability group* and reading and elementary)	all text fields	1987-present	140
What Works Clearinghouse	Small group reading	Literacy All Publications All Ratings	All years	19
Hand Search	*Elementary School Journal, Journal of Educational Psychology The Reading Teacher, Reading Research Quarterly, Review of Research in Education, and Journal of Literacy Research*	Tables of Contents Title and abstract	1994-2014	81
Total # of hits searching all fields: 35,584 Total # of hits searching abstracts/titles: 462 studies Total after duplicates were removed & screened using inclusion criteria: 243 studies Total # of studies meeting inclusion criteria: 29 studies (23 with ES, and 6 without ES)				

Inclusion and Exclusion Criteria

Cooper, Hedges and Valentine (2009) suggest guidelines for selecting eligibility criteria that should include eligible research designs, key study features and outcomes, geographical/language limitations, and any time frame restrictions. Included studies in this research synthesis met eight criteria:

- *It must be an empirical study of small-group reading instruction in the general classroom.* Empirical study designs used experimental and control groups and reported outcomes measures. Quasi-experimental designs were included Literature reviews and meta-analyses were excluded. Studies that used secondary analysis of nationally representative data sets, e.g., ECLS-K, were included if other inclusion criteria were satisfied. Qualitative studies (e.g., case study, ethnography, brief/reports) and studies missing statistical data were not automatically excluded; if they met all other inclusion criteria, they were coded and set aside for separate analysis.
- *The study had undergone peer review and was published or available between 1987-2014.* The year 1987 was chosen as the start date because the first comprehensive review of elementary ability grouping for reading addressed all studies prior to 1987 (Slavin, 1987), and subsequent reviews have focused on within-class groups or older populations (Kulik & Kulik, 1992; Lou et al., 1996; Lou et al., 2000). Studies must be published in English and address instruction in English-speaking classrooms.
- *Study participants were elementary students in grades ranging from Kindergarten to fifth grade*. K-5 was selected as the age-range since teacher-directed small-group reading instruction in general education occurs most often in the elementary grades and is less frequent in the middles grades and beyond (Loveless, 2013).
- *The independent variable focused attention on small-group reading instruction in the general education setting.* Students received their primary or Tier 1 reading instruction in a regular classroom setting in ability-based groups directed by a classroom teacher, specialist, or other trained adult. The small-groups in the study met for at least four or more sessions across the study. Studies targeting reading interventions, single reading skills/strategies for enrichment, writing, and small groups used for content area

instruction were excluded.

- *Participants in the treatment or control conditions met in homogeneous reading groups.* Group membership criteria could be assigned based on instructional reading level, reading strategy, oral/silent fluency, or other assessed reading skill.
- *Small-group was defined as ranging in size from 2-10 students.* This number was consistent with existing studies of reading intervention small-group formats (Elbaum et al., 1999).
- *Studies must include at least one outcome measure of student reading achievement.* Studies that reported measures of student or teacher self-concept, attitudes, or classroom learning behaviors/motivation were excluded from this study.
- *Studies reported the outcome data in a format amenable to effect size calculation*, i.e., means and standard deviations of treatment and comparison groups or allowed conversion using *t*-, *F*-, or *p*- values.

Exclusion criteria. Studies involving mastery learning, cooperative learning, peer tutoring, and multi-age non-graded classrooms were excluded from the synthesis. Mastery learning is frequently organized for individualized instruction and does not fit the definition of a "group." Cooperative learning and non-graded classroom studies typically rely on heterogeneous small groups, which is not the focus of this study. Peer tutoring involves students in the roles of teacher and learner and is typically conducted in dyads. Finally, studies of readers' workshop and literature circles were excluded unless they employed homogeneous teacher-directed small-groups for reading instruction. Traditionally, readers' workshop and literature circles are conducted with whole classes or with heterogeneous peer-led small-groups.

Coding Procedures

To code studies systematically and reliably for later analysis, a coding guide was constructed using an Excel spreadsheet following guidelines described by Cooper (2010). Lou, et al. (1996) state, "The purpose of coding study features is to identify those methodological and substantive characteristics which may be responsible for significant variations in the findings" (p. 431). After the initial computer search, 243 studies were searched and compared initially with the inclusion criteria to select studies for analysis. Studies that met all but two of the inclusion criteria were then set aside for an additional full-text search and coding into an Excel spreadsheet. Studies with multiple outcomes or independent sub-samples were recorded in separate columns in the same row.

Coders training. To protect against bias during study selection, three volunteer coders provided a reliability check for the inclusion and coding of a randomly selected sample of the studies in order to refine the codes, and to protect against "coder drift." The reviewer talked with the coders prior to training to discuss and refine the inclusion criteria and categories of codes. During the initial screening and final coding, the reviewer selected a random sample of 15 studies to check for inter-rater reliability and to spot-check for "coder drift." All studies were assigned a numerical identity code. A random number generator was used to select the studies for the coders. Coders were assigned a minimum of five studies each. They decided if the studies met the inclusion/exclusion criteria, and coded them into a separate Excel spreadsheet.

To train the coders, the reviewer provided an hour-long training session that consisted of an explanation of the project and research questions; a description of the coder's role; a review of each item in the coding manual; and modeling and practice using the coding guide on examples of excluded and included studies. Coders were instructed to leave the field blank if the information was not applicable to the reviewed study or code "not reported" if the information

was missing from the study. They were encouraged to write down page numbers and tables as a guide back to the effect size information. The primary investigator discussed any questions coders raised about using the coding guide and provided some online resources on study design and statistics. At the training conclusion, each received a random set of five studies to read and code. Author names and affiliations were removed from the selected studies in order to minimize potential coder bias. Coders were encouraged to contact the reviewer to ask questions about the coding guide, study design, or features for coding.

The coders held master's or Ed.S. graduate degrees in reading education and were all certified reading specialists with extensive experience serving as coaches or literacy leadership in their respective divisions. All three had taught graduate level reading courses and designed professional learning experiences for adults, and between them had over 75 years of experience in teaching reading at the elementary level. All had a basic understanding of statistics and research design. Inter-rater agreement was calculated by comparing the number of agreed-on codes by the total number of codes available (n=1650) to insure unbiased and reliable retrieval of studies. Inter-rater agreement rates ranged from 81% to 96% with a mean 90% agreement rate. All coding discrepancies and questions were resolved through conversations during virtual "real-time" meetings, emails, and phone calls until a 95% agreement rate was reached.

Coding guide and manual. The coding and organizing of studies was an iterative process intended to be replicable. The coding guide was set up using Excel software and included information to gauge each study's design and relevant features. The coding guide and manual in Appendix A were based on guidelines found in Cooper (2010, pp. 85-101) and from prior published meta-analyses on ability grouping (Elbaum et al., 1999; Lou et al. 1996).

Stock (1994) suggests seven categories for organizing information about each study: 1) report identification, 2) setting, 3) subjects, 4) methodology, 5) treatment, 6) process, and 7)

effect size. The preliminary coding guide was constructed with categories designed to answer the research questions; however, during the initial screening of the 243 studies, it was necessary to add additional categories in order to capture more information about each study. Examples of categories that were added included "total number of small groups" and "total time allotted for reading instruction." Additional categories were added if at least one reviewed study addressed the category and fit the study's research questions. No categories were removed.

Missing Data

Included studies were mined for effect size information by using study-provided effects or by recording the statistics needed for effect size calculation. Information for effect sizes was missing from 21% (n=6) of the included studies. All studies with missing data had been published prior to 2001. Every attempt was made to gather the data within the study, but if means and standard deviations were unavailable, calculations of effects were estimated by converting other statistics, e.g., *t*-values, *p*-values, and *F*-values. The primary reasons for missing data are due to inprecise reporting from researchers and looser publication standards for older studies. Although standard practice has changed to require the reporting of means, standard deviations, effect sizes, confidence intervals, and other descriptive statistics (Publications A.P.A., 2008), many studies published over ten years ago may not have followed this standard.

Ethical Concerns

Instead of using live subjects, empirical studies served as the subjects of the research synthesis. Because the authors for each study included in the proposed review presumably obtained permission from their respective institutions' ethical review boards, there are no "live" participants, and, therefore this removes many of the ethical issues that arise when conducting research with live participants.

Study Design and Rigor

The results of a research synthesis hinge upon the quality of the research designs included for analysis. The gold standard for research designs are true experiments that employ random selection and assignment to carefully measure and control the variables under study in order to demonstrate causal relationships. Unfortunately, true experiments are less common in educational research due to ethical, legal, and practical concerns (Cook, 2002; Schneider, Carnoy, Kilpatrick, Schmidt, & Shavelson, 2007). Empirical studies using experimental or quasi-experimental designs that employed treatment and comparison or control groups were sought for inclusion in the review, understanding the limits it places on reporting causal findings.

Because the included studies had widely varying research designs, a set of seven design and quality criteria was applied to each study and is listed here as a guide for readers of this synthesis. Using Elbaum et al. (1999) as a guide, a design quality score was assigned to each of these questions based on the following scale:

(a) presence or absence of experimental and control group (0-1pts);

(b) use of random selection and random assignment (2 pts), using non-random methods or matched groups (1 pt), or comparisons not used or unreported (0 pts);

(c) reports comparability of treatment and comparison/control groups (1 pt) or information not given (0 pts);

(d) provides strong evidence of a fidelity check for independent variables (2 pts), provides weak evidence of a fidelity check (1pts); or no evidence for fidelity checks is found (0 pts)

(e) unpublished dissertation/ MA thesis (0 pts); report available without review process, conference paper, policy brief, ERIC document, etc. (1 pt), or published with peer review (2pts).

(f) study design is a randomized experiment or quasi-experimental design (2 pts);

qualitative design with safeguards and quality standards (1 pt); not a study (0 pts)

(g) based on a small sample (<20) (0 pts); medium sample (21-100) (1 pt), large sample (>101 participants) (2pts)

Scores on these criteria were totaled to provide the reader with a rough estimate of the reviewed studies' quality with scores closer to zero suggesting weaker designs/less rigor and scores closer to twelve suggesting stronger designs. See Appendix A for a complete list of included study features and designs.

Unpublished and qualitative studies were not automatically excluded but were set aside to provide narrative description and details to support the review's central findings. Nobit and Hare (1988) describe meta-ethnography, or methods akin to meta-analysis used by qualitative researchers to create reasonable interpretations across multiple qualitative studies. They suggest organizing qualitative studies by common themes, concepts, or metaphors and interpreting the common "reciprocal relationships" (p. 38) among studies.

Validity and Reliability

The conclusions drawn from a research synthesis are worthless unless the review can demonstrate that measures were enacted to protect the data and results from potential bias and errors of interpretation. Analogous to a primary research design, the research synthesis must demonstrate trustworthiness by addressing concerns about reliability and validity. Following study quality guidelines outlined for Cochrane systematic reviews (Higgins & Altman, 2008), this study protects against threats to bias by addressing the following:

- external validity, i.e., does the study answer the research questions and are these questions important purpose for the educational community? Information in the first two chapters establish the relevance and significance of the research questions. Rigorous search and coding procedures and study inclusion criteria were applied to insure external

validity.

- internal validity, i.e., is the method of the research synthesis free from bias and therefore trustworthy? The risk of bias decreases by using multiple coders, providing the search terms and including clear inclusion/exclusion criteria.
- reliability, i.e., are the method and results of this synthesis consistent and replicable by others? The search and coding process has been carefully detailed and explained so that other researchers may attempt to replicate these results.

Table 5 provides a summary of the threats and steps taken to insure trustworthiness during the literature search, study selection, coding procedures, analysis of selected studies, and reporting of results.

Table 5

Summary of Threats and Protections from Bias

Bias	Threat to...	Protections within the Study
Does the study's results answer the research questions? Are the results relevant?	External validity	Research questions are clearly stated and linked to the discussion in Ch. 4. Ch 1 & 2 provide the statement of the problem and its relevance to the educational community. Number and type of citations examined for relevance.
Bias in the selected studies for analysis, e.g., missing key studies or missing data)	Internal validity	Representative sample by searching both published & unpublished sources Detailed record of search terms (replicable) Rate selected studies according to design criteria (Elbaum et al, 1999)
Bias in the selected studies, e.g., including irrelevant studies	Internal validity	Coding Manual Training additional coders & checks for inter-rater reliability
Search and results are not clearly explained	Reliability	Coding Manual Key Search terms provided Transparent search process and coding procedures so study is replicable

Failure to accurately account for the differences or lack of differences between selected studies.	Internal validity	Calculation of effect sizes creates a standard metric for comparing studies
Large studies with multiple outcomes and effect sizes may bias results.	Internal validity	Weighting effect sizes by calculating average effect sizes (Cooper, 2010) Random-effects so small studies are not discounted

Meta-analysis Plan

A quantitative and qualitative analysis was conducted of the 29 included studies. The quantitative analysis consisted of retrieving or calculating effect sizes (ES) from each quantitative study. Converting study outcomes to a common metric allows different research designs to be compared from a statistical perspective (Borenstein et al., 2009). First, effect sizes for reading achievement outcomes for each study were directly entered or calculated using the methods described below. Effect sizes measure the magnitude and direction of the group differences by converting group pre- and post- outcome differences to a standardized scale-free score. This allowed direct meaningful comparisons of the included studies' outcomes.

Several studies reported more than one measure of reading achievement or comparisons and reported multiple effect sizes within a single study. Since multiple effect sizes obtained from the same sub-sample of students are likely to be related, this could lead to misinterpretation of the study's overall effects and lead to Type 1 errors (Lipsey & Wilson, 2001). To obtain a mean overall effect for each independent study, three steps were taken: (a) if more than one reading outcome within the study purported to measure the same underlying construct, the effect sizes were coded for each, averaged, and a mean effect size was recorded; (b) all of a study's effect sizes were entered and analyzed if it was clear that the outcomes came from independent samples within the study. This was used if the study authors reported that there were independent groups

of participants, e.g., a second study within a study; (c) if the study reported multiple components of reading but the subjects were not from independent samples, then a single effect size was obtained via an *a priori* decision to select the outcomes that measured either overall comprehension or global reading achievement with a standardized instrument that demonstrated proven reliability and validity checks.

For studies that did not report effects, formulas were applied following procedures described by Lipsey and Wilson (2001) for *t*-values and *F*-values. Study statistics were entered directly into the Comprehension Meta-Analysis Software (CMA), which then computed the effect sizes automatically using Hedges *g* as a protection for small sample sizes (Borenstein, Hedges, Higgins, & Rothstein, 2005). As a check for accuracy, an online calculator designed by Wilson (n.d.) was used to confirm reported and calculated effects. If insufficient or missing data prohibited the calculation of effect sizes, then studies were set aside and reviewed afterwards using traditional qualitative methods (Nobit & Hare, 1988).

After calculating overall effect sizes for each study, a meta-analysis was run using the following steps: (a) estimated the distribution of the included studies' effect sizes and determined an overall effect size mean; (b) established a confidence interval around the mean effect size; (c) tested whether the effect size mean differed significantly from zero; and (d) used a test of homogeneity to examine the consistency of reported outcomes within the studies, i.e., do the included studies reflect the same (homogeneous) or multiple results.

Since the number of studies that met all inclusion criteria was small (k=23), and study sample sizes varied widely ranging from 4 to over 10,000, $\overline{N}$ = 2,251), the reviewer did not run an analysis for potential moderator or mediating variables. In situations with a handful of studies, Valentine, Piggott, and Rothstein (2010) suggest reviewers describe the characteristics and effect sizes and restrict their discussions to the current limitations of the research base.

Fixed- versus Random-effects Models

Prior to the meta-analysis, the reviewer decided whether to analyze the effects based on a fixed-effects or random-effects model. The choice of model influences whether the reviewer seeks to describe a "true" effect size (fixed) or to describe an overall distribution of effects (random). The formulas for each model are basically the same, except in how they deal with the sampling error found in the included studies. A fixed-effects model assumes that all of the included studies reflect a single, common effect, and the only reason that they find different effect sizes is due to sampling errors within each study (Borenstein et al., 2009). Fixed-effects models calculate an overall or "true" effect size mean by adding additional weight to studies with larger sample sizes and assigning their effects greater weight when calculating the effect size distribution.

If a fixed-effects model is used, it would assume that all of the studies included in the synthesis measured the effect of ability grouping on reading and, therefore the results could be generalized to populations beyond those studies (Borenstein et al., 2009). When using a fixed-effects model, greater weight is assigned to the outcomes of studies with larger sample sizes and more statistical power, since it is assumed that they produced a better estimate of the "true" effect size.

By contrast, random-effects models estimate a larger distribution of the overall effects of ability grouping on reading instruction, because the formulas take into account the additional variance that occurs both **within** and **between** studies. Using a random-effects model, the relative weights assigned to large and small studies are more balanced so that the effects of larger studies become less pronounced, and the contributions of smaller studies are not lost. Therefore, random-effects models tend to be less precise and do not allow for generalization to the population beyond the included studies in the synthesis.

For this synthesis, we can assume that the included studies had students who were provided varying amounts and types of small-group instruction under different conditions and used a variety of reading measures. Since they are unlikely to represent a single population, the synthesis used a random-effects model. In situations where there is no between-study variability, random-effects models return the exact same results as a fixed-effects model, so most experts recommend using a random-effects model first to account for as many sources of error as possible (Bornstein et al., 2010, 2011; Lipsey & Wilson, 2001).

Effect Sizes (ES)

The primary unit of interest in a research synthesis is the effect size (ES). Effect sizes provide a "common scale" to compare the outcomes of studies (Glass 2006, p. 429). Effect sizes provide a standard metric of the magnitude of difference between two groups and allow comparisons of different outcome measures as long as they are measuring the same underlying construct (Cooper, 2009, p. 164). For studies where means and standard deviations were not reported, statistical results were converted to effect sizes using formulas found in Lipsey & Wilson (2001).

Since all reading outcomes coded in the studies reported using an interval or continuous scale, standardized mean difference effect sizes were calculated were calculated by subtracting the experimental group's mean performance from the control group's mean performance and dividing by the control's standard deviation to yield a *d*-index (Cooper, 2010):

$$d = \frac{\bar{X}_1 - \bar{X}_2}{SD_{within}}$$

where $\bar{X}_1$ and $\bar{X}_2$ = the two group means; and SD_{within} = the estimated common standard deviation of the two groups.

To estimate the SD_{within} this formula was used:

$$SD_{within}\sqrt{\frac{(n_1-1)SD_1^2+(n_2-1)SD_2^2}{n_1+n_2-2}}$$

where SD_1 and SD_2 = the standard deviations of Group X_1 and X_2, and n_1 and n_2 = the sample sizes in Group X_1 and Group X_2.

Effect sizes in studies with large sample sizes are assumed to have greater statistical power and less sampling error than effect sizes from studies with small sample sizes; therefore the effect size should be weighted by the inverse of the sampling error variance (Lipsey & Wilson, 2001). Given that many of the included studies had small sample sizes (N <30), Hedges *g* (1981) based on standardized mean differences was calculated to provide more precision in determining the effect size and to minimize any bias caused by studies with small sample sizes. Hedges *g* was calculated using the formulas provided by the CMA software found in Borenstein et al. (2009):

$$g = \left[1 - \frac{3}{4n-9}\right] * ES_{smd}$$

where g = *adjusted ES*, n= sample size and ES_{smd} = standardized mean difference.

Missing means and standard deviations. When means and standard deviations were unreported, effect size was calculated based on *t*-, or *F*-values if the direction of the mean difference of the relationship was known using formulas developed by Rosenthal (1984) and described by Cooper (2010):

$$d = \frac{2t}{\sqrt{df_{error}}}$$

where t = the value of the t-test for the comparison, and df_{error} = the degrees of freedom associated with the error term of the t-test (n_1+n_2-2).

$$d = \frac{2\sqrt{F}}{\sqrt{df_{(error)}}}$$

where F = the value of the F-test for the comparison, and df_{error} = the degrees of freedom associated with the error term of the t-test $(n_1 + n_2 - 2)$.

Power Analysis

Once all effect sizes were recorded, a power analysis was used to determine if a sufficient number of studies warranted the use of meta-analysis. Since the synthesis used stringent conditions that limited the number and scope of independent variables, i.e., grouping and small group format, and dependent variable, i.e., reading achievement, then relatively small variability should be expected across the study outcomes. This was assessed using the I^2 index described in Valentine et al. 2010, p. 222. The I^2 index applied the following formula:

$$I^2 = \frac{Q-(k-1)}{Q} \text{ X } 100$$

I^2 values can range from 0 – 100%, and variability among study outcomes can be interpreted as ranging from 25% (small), 50% (medium), or 75% (large). The initial comparison of all of the studies suggested extremely large variability, $I^2 = 97.083$ and therefore a random-effects model was implemented, and subsequent moderator analysis was not implemented.

Test of Homogeneity

An overall test of the homogeneity of the mean effect sizes for the included studies indicated whether the differences between the distribution of effect sizes were due to chance, sampling error, or moderating variables within the studies. The homogeneity of the ES was measured using a Q statistic with the formula described by Hedges and Olkin (1985), where Q is distributed as a chi-square with $k - 1$ degrees of freedom and k represents the number of effect sizes:

$$Q = \left(\sum w_i ES_i^2\right) - \frac{\left(\sum w_i ES_i\right)^2}{\sum w_i}$$

where ES_i = individual effect size and $i = 1$ to k; k = the number of effect sizes, and w_i = the individual weight for ES_i (Lipsey & Wilson, 2001).

During the initial analysis of the reviewed studies, an effect size distribution was calculated based on the weighted means of the reviewed studies, a significance test was performed, and a test for the heterogeneity was conducted. The test for heterogeneity produced a significant Q-value, $p < .000$, suggesting that the studies' main effects varied more widely than expected by chance and represented random differences within and between the studies that cannot be identified. Subsequently, all remaining results were analyzed using a random-effects approach to account for any sampling error that is found across and between each reviewed study (Borenstein, Hedges, & Rothstein, 2007; Lipsey & Wilson, 2010). The goal of subsequent analysis was to describe the mean distribution of effect sizes for the reviewed studies.

Because of wide-ranging sample sizes and the small number of reviewed studies overall, a sensitivity analysis was used to check for the disproportionate influence of any single study on the outcomes. Systematically removing one study at each calculation did not make any significant change in the effect-size distribution, so all studies were left in the analysis.

Qualitative Analysis

Studies that met all inclusion criteria but did not allow for the calculation of effect sizes were not automatically excluded from this synthesis. Instead, qualitative analysis techniques were applied following the steps of a traditional narrative review (Pope, Mays, & Popay, 2007). There were six studies that were missing data needed for effect size calculation or were qualitative studies of small group instruction. The procedures to analyze the findings of these studies

involved providing descriptions of the studies' sources, identifying their major themes, and presenting their key findings.

Summary

Chapter 3 provided a step-by-step explanation of the procedures used during this research synthesis beginning with an exhaustive literature search for the studies focused on within or between class ability-grouping for reading instruction in the general education classroom. Study inclusion and exclusion criteria were applied using a coding guide and consistent procedures. Additional coders were used to decrease threats of selection bias and insure accuracy of coding. All initial coding was done using Excel software. Effect sizes were calculated and analyzed using CMA software and an online ES calculator (Wilson, n.d.). Care was taken to obtain overall mean effect size for each independent sample within a study on reading outcome measures. Ethical concerns and protections against bias were addressed throughout the search, coding process, and analysis. If study effect sizes were unavailable, then studies were set aside for qualitative review. Chapter 4 will describe the results of the synthesis, and Chapter 5 will discuss the implications of the findings.

CHAPTER 4

FINDINGS OF THE STUDY

The purpose of this research synthesis is to explore the use of ability grouping in elementary reading instruction and to describe empirical support for the elements and formats of small-group reading instruction in the elementary general classroom. Chapter 4 describes the results of the synthesis for each comparison in the analysis.

Interpreting Effect Sizes

The primary unit of interest in a research synthesis is the effect size. Effect sizes are a standard measure of the difference and magnitude of difference between two groups. Effect sizes yield a result that ranges from 0 to 1 and can be interpreted like a standard *z* score or a standard deviation score. Zero indicates no difference between the groups; scores closer to 1 indicate differences between the groups (Lloyd, Forness, & Kavale, 1998). Although interpretation of the magnitude of effect sizes can vary, educational researchers generally accept the following: small change (<.50), medium change (.50-.79), large change (>.80) (Cohen, 1988).). In *Visible Learning,* Hattie (2008) synthesized over 800 meta-analyses of studies on student achievement and found the average effect size for most educational studies was .40, therefore, .40 was suggested as a "hinge point" to determine what works "best" in education.

Effect sizes in educational studies are rarely more than .40, so effect sizes over .40 are considered to demonstrate a meaningful change in student performance (Hattie, 2008; Lloyd et al., 1998).

Characteristics of Included Studies

Twenty-nine studies met the inclusion criteria for the synthesis with 23 studies that

reported a total of 68 effect sizes. The remaining six studies are reviewed separately at the end of this chapter. All 29 reviewed studies were published in peer-reviewed sources, i.e., journals, chapters, or books, and 25 employed quasi-experimental or experimental designs. Appendix B provides a list of all included studies with information about the type of ability grouping, i.e., within- or between-class groups, format of small groups in each study, grade, reading outcomes, i.e., reviewed studies' dependent measures, sample and effect sizes, and study design/rigor. Excluded studies (n=214) along with the reason/s for exclusion may be found in Appendix C.

Quantitative Analysis of Study Outcomes

As discussed in Chapter 3, the initial analysis used a random-effects model to provide a less biased estimate of the effect size distribution by accounting for additional variance between studies. The overall comparison of the effects of elementary ability grouping for reading instruction ($n = 23$) indicated a small but significant effect supporting the influence of ability grouping on reading achievement with a mean effect size of +.30, ($p < .0001$), and a 95% confidence interval ranging from +.17 to +.43. Previous reviews of ability grouping have reported similar findings from 19-56 studies using similar inclusion criteria (Elbaum et al., 1999; Kulik & Kulik, 1992: Lou et al., 1996; Slavin, 1987). The number of included studies fell within the low-end of this range; however the inclusion criteria for this synthesis targeted K-5 students in the general classroom setting.

The initial comparison included both types of homogeneous groups: within-class and between-class versus non-grouped classrooms. See Table 6 for a list of included studies and their characteristics. The remainder of the chapter will discuss the remaining comparisons of the type of grouping, reading outcomes, grade/age, and end with a discussion of the findings on instruction within the small groups.

Table 6

Overall Ability-Grouping vs. Non-Grouped Classes

Study	Hedges g	*Lower Limit*	*Upper Limit*	*p-value*
1. Adelson & Carpenter_2011	0.92	-0.23	2.07	0.12
2. Borman, Slavin, Cheung et al._2005	0.22	0.13	0.30	0.00
3. Buttaro & Catsambis et al_2010	0.39	-0.52	1.30	0.41
4. Chambers Abrami et al. 1998	0.17	-0.08	0.41	0.18
5. Condron_2008	-0.10	-0.15	-0.04	0.00
6. Greenlee & Bruner_2001	-0.88	-1.62	-0.14	0.02
7. Hanselman & Borman_2013	0.27	0.16	0.39	0.00
8. Hong & Hong_2010	0.24	0.08	0.38	0.00
9. Hong_ 2010	0.16	0.00	0.32	0.05
10. Hong, Corter et al_2012	0.27	0.05	0.49	0.02
11. Kamil & Rauscher_1990	1.61	1.43	1.79	0.00
12. Lleras & Rangel_2009	-0.16	-0.25	-0.06	0.00
13. Madden, Slavin, & Karweit_1993	0.69	0.21	1.17	0.01
14. McCoach, OConnell & Leavitt_2006	0.81	0.37	1.24	0.00
15. Robinson_2008	0.17	0.02	0.32	0.03
16. Ross & Smith_1994	0.09	-0.32	0.49	0.67
17. Ross Smith & Casey_1999	0.13	-0.13	0.38	0.33
18. Ross, Nunnery et al._2004	0.12	0.11	0.13	0.00
19. Scarcelli & Morgan_1999	0.23	-0.29	0.76	0.39
20. Slavin Madden Karweit et al_1990	0.49	0.12	0.85	0.01
21. Smith Ross & Casey_1996	0.50	0.06	0.94	0.03
22. Tach & Farkas_2006	0.60	0.55	0.65	0.00
23. Tivnan & Hemphill_2005	0.41	0.24	0.57	0.00
Random-Effects Model	.299	.16	0.43	0.00

Analysis of Within Class Ability Groups (n=13)

Thirteen studies reported findings on the use of ability groups for reading, and of those studies, the majority (n=11) used data from the U.S. Department of Education's Early Childhood Longitudinal Study program (ECLS), a series of three longitudinal studies of early child development and school experiences. Beginning in 1998-99, the ECLS-Kindergarten (ECLS-K) followed a cohort of kindergarten through eighth grade students. Since the first release of the ECLS-K data in 2001, it has created opportunities for researchers to probe early school experiences using a large nationally-representative sample (McCoach, O'Connell, & Leavitt, 2006). One of the more robust findings from this part of the analysis was a significant positive effect, (ES = +.373, $p < .001$) supporting the use of within-class ability groups for reading versus whole group instruction. See Table 7 for a list of studies. Although generally positive effects were reported for grouped versus non-grouped students (Hong, 2010; McCoach, O'Connell, & Leavitt, 2006; Tach & Farkas, 2006), most (n=11) reported varying effects for children placed in low, middle and high groups for reading instruction (Adelson & Carpenter, 2011; Hong, Corter, Hong, & Pelletier, 2012), with three studies publishing either negative or non-significant outcomes for students assigned to the lowest groups (Condron, 2008, Kamil & Rauscher, 1990; Lleras & Rangel, 2009). Notably, three of the studies conducted by Hong and Others (Hong, 201, Hong & Hong, 2010, Hong et al., 2012), reported that children in the low groups either out-performed non-grouped students when grouping occurred under high-intensity conditions, i.e., teachers spent more than one hour daily on language arts and devoted at least 40% or more of their total language arts block to small group instruction.

Five of the within-class studies using homogeneous ability grouping examined the impact of ability grouped versus non-grouped classrooms on minorities and ELL students, specifically Hispanic and African Americans (Buttaro & Catsambis, 2010; Condron, 2008; Lleras & Rangel,

2009; Robinson, 2008). Buttaro & Catsambis (2010) focused on cultural, school politics, and organizational factors that influence ability group placement and outcomes and found an overall positive influence from grouping. The studies were evenly split with two reporting overall positive effect sizes for minority students (Buttaro & Catsambis, 2010; Robinson, 2008) and two studies that did not (Condron, 2008; Lleras & Rangel, 2009). Tach and Farkas (2006) found no differences in ability group placements for African American students compared to White students after controlling for classroom learning behaviors.

Table 7

Comparison of Within-Class Ability Groups vs. Non-Grouped Classes

Study	Hedges g	*Lower Limit*	*Upper Limit*	*p-value*
1. Adelson & Carpenter_2011	1.34	0.04	2.64	0.04
2. Buttaro & Catsambis et al_2010	0.39	-0.52	1.30	0.41
3. Condron_2008	-0.09	-0.15	-0.04	0.002
4. Hong & Corter et al._2012b	0.16	0.00	0.32	0.05
5. Hong, Corter et al_2012	0.27	.002	0.32	0.05
6. Hong & Hong_2010	0.24	0.08	0.40	0.003
7. Kamil & Rauscher_1990	-1.61	1.43	1.79	0.000
8. Lleras & Rangel_2009	-0.16	-0.25	-0.06	0.002
9. McCoach, OConnell & Leavitt_2006	0.81	0.37	1.24	0.000
10. Robinson_2008	0.17	0.02	0.32	0.023
11. Scarcelli & Morgan_1999	0.23	-0.29	0.76	0.000
12. Tivnan & Hemphill_2005	0.54	0.30	0.78	0.000
13. Tach & Farkas_2006	0.60	0.55	0.65	0.000
Random-Effects Model	.373	.191	.555	.000

Analysis of Between-Class Ability Groups (n=10)

The ten studies in this comparison of the synthesis were all evaluations of the nationally

recognized school reform program called Success For All (Slavin et al., 1990) described in Chapter 2. The What Works Clearinghouse (2007, 2009) has issued two intervention reports on the quality of SFA, with the latest review finding seven studies that reported positive and mixed mean effects for SFA that ranged from +8 to +13 percentage points. None of the evaluations of the SFA program have teased apart the specific contributions of its individual programmatic elements, i.e., cross-grade grouping, to the overall success of the model. See Table 8 for a list of studies.

Because all ten of the studies of between-class ability grouping included in this synthesis were evaluations of SFA, and SFA does not implement cross-grade groupings until 1st grade, scores reported for kindergarten students were not included in the overall effect size calculations for this synthesis. This may create potential differences between the effect sizes reported in this synthesis and other reviews. For this synthesis, the ten included studies of between class grouping yielded small but significant mean effects for students beyond kindergarten with a mean *ES* of .205, $p < .001$, with a 95% confidence interval between .09 to .33. All of the included studies used experimental (n=1) or quasi-experimental (n=9) designs with safeguards for study quality. Three of the ten studies were conducted by third-party evaluators (Chambers, Abrami, Massue, & Morrison, 1998; Greenlee & Bruner, 2001; Smith, Ross, & Casey, 1996), and the rest were conducted by SFA researchers (n=4) or their funding affiliation was not reported (n=3). Since between-class ability grouping is confounded with the other SFA program factors, it is impossible to tease apart the impact of between-class grouping on small-group reading instruction based on these studies.

Table 8

Comparison of Between Class Ability Groups vs. Non-grouped Classes

Study	Hedges g	*Lower Limit*	*Upper Limit*	*p-value*
1. Borman, Slavin, Cheung et al._2005	0.22	0.13	0.30	0.00
2. Chambers Abrami et al. 1998	0.17	-0.08	0.41	0.18
3. Greenlee & Bruner_2001	-0.88	-1.62	-0.14	0.02
4. Hanselman & Borman_2013	0.02	-0.23	0.27	0.88
5. Madden, Slavin, & Karweit_1993	0.69	0.21	1.17	0.01
6. Ross & Smith_1994	0.09	-0.32	0.49	0.67
7. Ross Smith & Casey_1999	0.13	-0.13	0.38	0.33
8. Slavin Madden Karweit et al_1990	0.60	0.24	0.97	0.00
9. Smith Ross & Casey_1996	0.50	0.06	0.94	0.03
10. Tivnan & Hemphill_2005	0.28	0.04	0.51	0.02
Random-Effects Model	0.21	0.09	0.33	0.001

Note: Since SFA does not use cross-grade grouping until 1st grade, reported scores for their kindergarten cohorts were not included in the overall effect size means.

Analysis by Reading Outcomes

At the heart of the synthesis is determining whether any impact of small-group reading instruction exists, and if so, is it more or less useful for specific reading outcomes. This section reports the comparison of the effects of small-group instruction according to children's reading outcomes. All 29 included studies had one or more outcome measures for reading. In cases with fewer than two studies reporting effect sizes for a category, no comparison was done. Outcomes from standardized achievement tests or researcher-created tests with fidelity measures were included in the synthesis. See Table 9 for a summary.

Table 9

Comparison of Reading Measures

Category (description)	# of studies	Mean ES	*p-value*	95% Confidence Interval
Alphabet / Letter Sounds/Phonics+ (measures of letter names, sounds, and production or spelling)	1	--	--	--
Comprehension (measures of understanding of passages after listening &/or oral/silent reading, cloze procedures)	3	.83	.181	-.39 to 2.04
Fluency of Oral Reading (includes measures using passage reading, timed passages, and informal reading inventories to assess reading level)	4	.25	.003**	.09 to .42
Global Reading (standardized achievement test or combined measures that result in a single score reflecting students' overall reading achievement)	18	.25	.03*	.02 to .48
Phonological Awareness (measures of phonological processing, e.g., segmenting and blending oral sounds in words	0			
Vocabulary (measures children's knowledge of words through expressive or receptive measures)	2	.84	.16	-.32 to 1.99
Word Identification/Word Attack (measures of accurate word identification typically with graded word lists or nonsense word lists)	6	.32	.001**	.13 to .51

Notes: +A minimum of 2 studies was required to calculate a comparison. * = $p<.05$ **= $p<.01$

All seven categories of reading outcomes were analyzed, but only the significant effects are discussed. Measures of alphabet /letter sounds were used frequently in studies, but often it was reported as a combined score with other batteries, e.g., phonics or word identification or relied on informal measures lacking fidelity checks. All of the outcome measures used continuous or interval scales to measure student reading.

Due to the small sample size, these comparison results should not be generalized to a broader population outside the included studies. However within the confines of the included studies, small-group reading instruction in the K-5 classroom yielded small but significant benefits on three of the outcome measures in order of magnitude: Word identification/Word attack (+.32), Fluency of passage reading (+25)., and Global reading achievement (+.25). Alphabet and Phonological Awareness are well-documented predictors of word identification and reading achievement; there were no included studies that used these as stand-alone measures of effects. It is worth noting that these components were combined and included to create a global achievement score (McCoach et al., 2006; Robinson, 2008) or used but there was insufficient data to calculate effects (Juel, 1990).

Analysis of Small-group Instruction Formats

The different formats used for small-group reading instruction in the regular K-5 classroom across the reviewed studies were examined. Each of the formats was described in Chapter 2, and Table 10 lists the small-group formats and their effect sizes. The main result from this comparison is that the literature search revealed a paucity of empirical studies on the use and efficacy on seven out of eight common formats used for general classroom reading instruction. Of the 23 studies in the ES analysis, the lion's share was divided evenly between studies reporting the ECLS-K data (n=10) and the SFA evaluation studies (n=10), with four additional studies representing Basal Reader, Differentiated instruction, and Four Block. It is important to note that one format, Interactive Strategy Approach (Vellutino and Scanlon, 2002) and its accompanying studies (n=6), was removed from the research synthesis prior to entry into the CMA software, since all of the studies addressed intervention or used 1:1 instruction; the focus was not general classroom small group instruction.

The SFA studies (n=10) reported a small but significant ES of +.20. As discussed earlier

in the chapter, SFA is a comprehensive school reform package, so it would be misleading to describe their effects as representing a specific instructional format since SFA has many layers of intervention and family supports that likely contribute to students' success.

None of the studies using ECLS-K data reported the type of format that the teacher used for small group instruction, and therefore, none were included in this comparison. Studies using the ECLS-K data reported general information about the amount of time and allocation of the ability groups, i.e., average number of groups, number of times groups met weekly, and the total amount of time allotted for language arts instruction. Little or no information was given to describe the microstructure, routines, or content of small group instruction. There is no way to discern whether teachers differentiated instruction within each group or taught the same texts and activities regardless of group membership. The independent variable in the ten studies, ability group reading placement, was based on teacher reportings of student rankings (McCoach et al., 2006), student characteristics (gender, race/ethnicity, social class) (Tachs & Farkas, 2006), or combining predictive factors to create artificial classes through complex statistical procedures, e.g., marginal mean weighting through stratification (Condron, 2008; Hong & Hong, 2010). Given the widely different schools, classes, and students in the ECLS-K data-set, it is unlikely that the students in these national studies shared a common small group instructional format across the various sites, but many may have experienced similar grouping experiences within a single school since teachers report that administrators direct them to use specific grouping formats (Moody et al., 1997).

Table 10

Comparison of Small-Group Instructional Formats

Category (description)	# of studies	Mean ES	*p-value*	95% Confidence

				Interval
Basal Reader Group	1			
Differentiated Small Group	1			
Four Blocks	1			
Guided Reading +	1	.54	n.s.	
Interactive Strategy Approach++	0			
PALS Electronic Lesson Plans	0			
Reader's Café/Daily Five	0			
Success For All Wings	10	.20	.001**	.09 -.33

Notes: ES = effect size * = $p<.05$ **= $p<.01$
A minimum of 2 studies was required to calculate a comparison. Blank columns indicate that no ES were calculated
+ Tivnan & Hemphill (2005) compared 4 formats but n.s. differences were found among the programs.
++This category was removed from the synthesis because the 4 studies found related to ISA all addressed intervention or used 1:1, not general small group instruction.

Analysis By Age/Grade

The final comparison of the meta-analysis examined the effects of grouping by the grade/age of students grouped to see if there were any significant differences for primary-aged students, defined as K-2 versus older students, defined as grades three to five (Table 11). Twenty studies yielded a mean ES for overall reading achievement in K-2 with ES = +.23, $p<.0001$. For the eight studies reporting effect sizes for reading groups for students in grades 3-5, the effects were not significant. This suggests that ability grouping had a small positive impact on the overall reading achievement of younger students.

Table 11

Comparison of Outcomes by Grade

Grade	# of studies	Mean ES	*p-value*	95% Confidence Interval

Kindergarten – 2nd grade	20	.23	.000	+.10 - +.34
Third – Fifth grade	8	.10	n.s.	-.08 - .29

Notes: * = $p < .05$ ** = $p < .01$ n.s. = not significant

Qualitative Results

There were six studies included in this review that did not allow for the calculation of effect sizes, but otherwise met all but one of the inclusion criteria for this study. Refer to Table 12 for a summary of the designs, type of grouping, and common themes and findings. A discussion of the studies and their common patterns and findings follows. Five of the six studies, with the exception of Henrickson & Jones (2013), were published prior to 1998.

Table 12

Qualitative Analysis of Remaining Studies

Study (n=6)	Design	Group Type	Grade	Themes	Findings
Anderson et al, 1991	Q-Exp	Within	3rd	Focus of instruction code versus meaning	Student performance can be predicted from group placement. Materials used during small groups may affect how well a group performs.
Cunningham, Hall, & Defee, 1998	Q-Exp	Within	1st	Response to ability grouping Shift to guided reading	A study describing the implementation of Four Blocks instruction in two 1st grade classrooms. Serious design flaws and insufficient data for ES. Note: change the basal group block to a guided reading block based on instructional reading level.
Dreeben & Barr, 1988	Descriptive	Within	1st	Grouping decisions & outcomes	Classroom setting, composition and allocation of resources affect teachers' decision-making to form ability-based small-groups for reading.
Hall & Cunningham, 1992	Qualitative	Within	1st	Response to ability grouping uses basal-reading groups as 1 rotation	Describes a rationale and approach to managing the language arts block in the regular classroom, called 4 Block. Seen as a non-ability grouped method, but uses within-class small groups based on

					basal reader for one rotation.
Juel, 1990	Descriptive	Within	1st	Outcomes of small groups and what materials are used	Describes how students are grouped and the types of materials used for 179 1st graders in 15 first & second grade classrooms.
Jones & Henricksen, 2013	Action-Research Descriptive	Within	1st	Skills-based ability groups Fluid, dynamic groups	Description of a 1st grade classroom and the implementation of skills-based small groups but there are no comparison groups, small sample (n =26). Argues for the use of differentiated skills-based reading groups in primary classes.

Dreeben & Barr (1988)'s landmark study of ability grouping decisions identified many of the sociological factors that impacted teachers' decisions about group placement. Using data collected from over 300 first graders in 13 classrooms, this was part of a longitudinal study of beginning reading. First and second grade achievement scores were compared to class distribution, group placement, i.e., high, middle, or low, pacing of basal instruction, and other learning design issues. Their results indicated that class composition shapes group placement and subsequently, reading achievement. One of the critical findings of this study was that classrooms with large distributions of low-ability students make significantly fewer gains in reading. Teachers must consider student individual differences in relation to how those differences are distributed across the class. The study suggests that "instruction is the mechanism that differs within small groups and how much instruction groups receive should influence how far they progress" (Dreeben & Barr, 1988, p. 138).

Juel (1990)'s study of 1st grade classrooms followed in the footsteps of Dreeben and Barr (1988). Also considered an important study of the use of ability groups, it provides a portrait of "how" students are assigned to groups and "what" types of materials are in use for 179 first graders in 15 first and second grade classrooms.

The descriptive studies of the use of Four Blocks instruction (Cunningham et al., 1998; Hall et al., 1991) were an attempt to shift from the use of ability-based reading groups in the first grade classroom. However, both the initial and follow-up studies reported using reading groups within the classroom for instruction. Without a comparison group and providing post-data only, the studies had major methodological flaws. It was included in the synthesis since it met the inclusion criteria, but insufficient data was unavailable to calculate an effect size.

Anderson, Wilkinson, & Mason (1991) conducted a mixed methods study that examined two formats for small-group reading instruction. Students were exposed to either a meaning-based or word-emphasis condition for four lessons. The same texts were used for all groups, regardless of instructional reading level. Although a counter-balanced design was used to vary the treatment conditions, the treatments were brief, two sessions each, and there were no fidelity measures for the treatment conditions. Although ability groups were used, no information was provided related to how the groups were formed, group size, or how often group membership changed. The low ability group had significantly lower fluency scores and difficulty reading the grade-level passages. Although this conclusion was not directly tested by the original study, the study's outcomes suggest that students of varying abilities benefit from the use of instructional level reading materials.

Jones and Henricksen (2013) described an action-research project in a single first grade classroom that illustrated the use of differentiated skills-based instruction for 26 students. As an empirical study it would have serious design flaws, but the standards for action research are looser. The study does not tell the reader many details about the students themselves in terms of race, ethnicity, and gender. Little information is provided about the children's pre- and post-test performance, and no comparison group or statistical analysis was used to evaluate student gains. The authors compare student outcomes on DIBELS to average national norms, but results are

shown in a bar chart format and lack specific numbers needed for effect size information. The other reported data is a comparison of the below-level and on/above level readers' post-test scores. Since we do not know whether the groups were equivalent or not at the beginning of the year (no pre-test data), this information may lull readers into thinking the small-group instruction produced significant gains. Based on the information provided, it is unclear whether the below-level readers made "real" gains or if their reported scores are simply due to chance.

Despite its statistical and design flaws, the study does provide some glimpses into the decision-making process that teachers use to plan small groups, and offers some description of the format that these lessons may take. They discuss ways to link assessment results directly to small-group reading instruction, but their described practice of coordinating six daily small groups in the classroom is impractical and unrealistic for most classroom teachers. If differentiated instruction results in the formation of six or more small groups, there is a trade-off between keeping group sizes manageable and meeting the most needs possible given allocated resources and personnel. For example, if a teacher coordinated six daily small-groups for 20 minutes each, that would mean two hours of small group time with all students spending 20 minutes with the teacher and a little over 1.5 hours in other activities, presumably independent work. It is unclear from this study and the others in the review, whether independent work contributes to reading outcomes.

What overall conclusions can be drawn from a qualitative examination of these six studies? All six provided a more nuanced approach to the inner-workings of small-group instruction in the regular classroom setting, i.e., the "how" and "what," but all but one of the studies (Henricksen & Jones, 2013) were published over ten years ago, which limits their relevancy to today's classrooms.

Four of the studies occurred during a public backlash against the use of ability groups in the classroom and may have represented the field's response to using undifferentiated whole group instruction (Dreeben & Bar, 1998; Cunningham et al., 1998; Juel, 1990; Hall & Cunningham, 1992). Anderson et al. (1991) with its "same text for everyone regardless of instructional level" format exemplifies the classroom conditions that pre-dated the call for differentiated instruction. This approach to instruction flies in the face of meeting children at their zone of proximal development using scaffolded activities and materials (Wood et al., 1976). Finally, Jones & Henrickson (2013)'s action-research study illustrated how the use of small groups for reading has evolved from a static, three-group (high-, mid-, low-) model using the same texts/materials for all (Anderson et al., 1991) to more dynamic patterns of grouping, i.e., flexible groups (Flood et al., 1992).

Summary

Chapter 4 reported the distribution of effect sizes found during the quantitative analysis of the included studies (n=23) and concluded with a qualitative summary of the six remaining studies. The overall findings show a small significant positive effect for the use of small-group reading instruction in the general K-5 classroom. The small number of available studies and wide variability within and between them limits any causal generalizations. The other critical finding from this analysis is the dearth of research for the formats, the "what" and "how", of K-5 ability-based small-group instruction for classrooms using between groups and also for commonly used instructional formats and routines.

The next chapter will interpret the scope and limitations of these results in terms of the synthesis' research questions. To avoid introducing error, procedural steps were taken at each phase to protect the search, selection, coding, and analysis to allow transparency and replicability. The final chapter will interpret these results to compare the quantitative and

qualitative findings, answer the three research questions, and finally, to discuss the limitations and implications.

CHAPTER 5

DISCUSSION

This research synthesis examined current practices for ability-based small-group reading instruction in the elementary grades by systematically searching for existing empirical studies, coding their relevant features, mining them for effect sizes, and using both quantitative and qualitative analysis procedures to describe the scope and limitations on the use of ability groups and formats for general small-group reading instruction. The dependent measure in this synthesis was reading achievement as measured by the included studies' measures of reading skills. The initial search and screening yielded a set of 243 studies that was subjected to full-text analysis. The final screening resulted in a set of 29 studies. Each study was mined for all related outcome measures that addressed the synthesis' dependent variable: student reading achievement. After effect sizes were calculated and organized using CMA software (Borenstein et al. 2011), a random-effects meta-analysis was conducted on 23 studies with k=68 effect sizes. Mean overall weighted effect sizes were calculated for each study in order to test their overall contribution to student reading outcomes. Tests of heterogeneity were significant at each comparison, suggesting that a random-effects model should be used to take into account any sampling error both within and between studies. A random-effects model suggests that the resulting effect sizes should be interpreted with caution, and conclusions limited to the population of studies in this review. Due to the small sample set and wide variability among the studies, moderator analysis was not used. Comparisons were made between the studies' weighted effect sizes to assess individually the contributions made by type of grouping, reading outcomes, age/grade, and format of small-group instruction. Qualitative analysis was applied to the remaining six studies

and compared to the earlier quantitative results.

In this chapter, the results from Chapter 4 will be used to interpret the results from both the quantitative and qualitative analyses and then use that information to address each research question. The chapter will discuss this synthesis' limitations and conclude with implications for future research, policy, and practice.

Comparing Quantitative and Qualitative Findings

Although direct comparisons cannot be drawn, there are some interesting patterns across the quantitative and qualitative analyses. First, the field's historical interest in the use of small-group reading instruction and studies examining the formats used for teaching small groups peaked during the late 1980's, faded during the 1990's, and may be rising again. Fourteen out of the 22 studies in the quantitative analysis (64%) were published after 2004, within the last 10 years, suggesting that the interest in small-group reading instruction in the general classroom is on the upswing.

Second the quantitative studies, though small in number (N=23), yielded effect sizes that suggest that small-group reading instruction within the general education classroom overall has small but positive overall impacts on reading achievement in the elementary grades (Adelson & Carpenter, 2011; Tachs & Farkas, McDonnell, Coach, & Levitt, 2006), and even more promising are the steady accumulation of studies that support the practice of homogeneous small-group instruction in the elementary grades to bridge the achievement gap for minorities, low SES (Hong & Hong, 2009; Hong et al., 2012, 2012a) and Hispanic/ELL students (Robinson, 2008). Still within this chorus, there are a few voices of dissent particularly around issues of race and SES and the long-term impact of reading failure (Condron, 2008; Lleras & Rangel, 2009).

Third, the quantitative comparisons of reading outcomes and age/grade comparisons hint that nuanced explorations of small-group reading instruction are needed to determine the scope and limitations of its impact on students.

Besides the comprehensive school reform of SFA, there were no studies on the use of between-class grouping in the elementary grades. Does this mean schools no longer use this practice or is it used but not reported? Currently, the reviewer knows of at least three schools in Virginia that currently use between-class grouping for reading instruction only (all non-SFA schools), so it is likely understudied.

This synthesis suggests that small-group reading instruction provides a small positive impact that is felt most in the primary grades, K-2 and fades to almost zero by grades 3-5. This finding for grades 3-5 was unexpected and surprising to the reviewer. Why should small-group reading instruction have less impact for readers in grades 3-5? It remains unanswered whether this is due to a decreased use of small-group instruction for older students (Loveless, 2013), whether it reflects a format that does not benefit older students, or reflects some other contributing factor, e.g., poor quality instruction. It is hypothesized to be some combination of the above. Schools commonly allocate less time to reading instruction for grades 3-5, e.g., 120 minutes for K-2 and 90 minute blocks for upper grades. If children are spending less time overall in language arts, this reduces the amount of time available for quality small-group instruction. Comparisons of the reading outcomes suggest that outcomes using word identification and word attack skills and fluent oral reading, skills that are mostly developed with younger K-2 children, were fostered using ability-based small group instruction.

It is well documented that students who fall behind in the primary grades have a tendency to stay behind throughout their schooling. Since there were so few studies that focused on grouping for students in grades 3-5, it would be misleading to spend time searching for

mediating effects until more primary research is done that targets the settings, situations, and nature and quality of small-group reading instruction that benefits older elementary 3-5 readers.

Research Questions

The findings from this research synthesis define the scope and limitations of what is known about the use of small-group reading instruction in K-5 classrooms. This section of the chapter will revisit each research question in light of the results and move from a broad perspective on the use of ability groups for reading in the K-5 classroom to a focused perspective on the formats of small group reading.

Research Question A

Since 1987, what evidence supports or refutes the use of ability-based reading groups in the general classroom in the elementary grades? The overall comparison of the effects of elementary ability-based reading groups yielded a small but significant effect, meaning that ability-based groups in the general classroom do exert a positive influence on students' overall reading achievement with a mean effect size of +.30, ($p < .0001$), and a 95% confidence interval ranging from +.17 to +.43. Conventions described by Cohen (1988) suggest a small to almost mid-sized effect of +.30 standard deviations. This means that using ability-based small groups for reading in the K-5 classroom resulted in a positive increase in students' overall reading achievement compared to children who were taught in non-grouped or other settings. Since the confidence interval does not include zero, we can be 95% sure that that the effect size is not zero. In practical terms, this suggests that students who received ability-based small-group instruction in general K-5 settings demonstrated small to mid-size increases in mean reading achievement over students in non-grouped regular classrooms. For the studies included in this synthesis, non-grouped referred to any classroom that used whole group or mixed-ability small groups for instruction.

Separate follow-up comparisons of the subset of studies using within-class and between-class grouping bolster this claim. Analysis of the mean effect sizes of studies using only within class grouping (n=13) increased to an overall mean effect of +.37, $p < .000$. The 95% confidence interval expanded for the within-class studies from .19 - .56, reinforcing that a small positive effect exists for students receiving ability-based small-group reading instruction that accounts for almost 27% of the differences between grouped and non-grouped students. The effects of between-class grouping will be interpreted separately, since it is the topic of the next research question, but the between-class studies (n=10) also lend support to the initial conclusion, albeit with a smaller overall mean effect of .20, $p < .0001$.

While the overall effects from each study suggested a positive influence, the effects of small-group instruction at different grades revealed an interesting difference between primary (K-2) and older elementary students (Gr. 3-5), namely that for the 20 studies reporting effect sizes for younger students (K-2), there was a small positive impact of +.23 of a standard deviation, but this effect virtually disappeared for older students in grades 3-5 (ES = +.10, n.s.). This finding suggests that small group reading instruction has gradually decreasing effects as students move through elementary school, i.e., it works better for younger students. Due to the low number of studies and effect sizes, these results may simply be due to chance and should be interpreted with extreme caution. An alternative explanation may be that primary grade teachers use small-group reading instruction less as students get older, a fact that is widely supported by the literature (Chorzempa & Graham, 2006; Loveless, 2013). More primary research is needed to substantiate and explore this finding.

It is also important to note that 10 out of the 13 within-class studies of this review were based on ECLS-K data that was gathered during the 1998-1999 school-year, a time period marked by a widespread use of whole-group instruction; the remaining studies in the full sample

(n=23) represented data gathered from students across 1989-2010. The students in these studies represent a wide-ranging population, and due to the nature of the data set, these studies used teacher self-reports to provide vague over-arching descriptions of the students' small-group reading instruction. Little is known about the type and quality of the instruction that these children received in their small groups.

The ECLS-K studies reflect data that was collected fifteen years ago and while it may give insight into the forces that shape early school experiences, it cannot substitute for current primary research. As more data is released from the ECLS studies, additional studies of the effects of ability grouping for reading with older elementary students will become available. The general descriptions of small-group instruction that these studies offer should spur interest in primary studies on the specific formats and outcomes of small group instruction, analogous to the way that the Response to Intervention movement has added momentum to the interest in small-group reading interventions (Gersten et al., 2008).

Research Question B

What evidence supports or refutes the use of between-class ability grouping, e.g., Joplin-type plan, for classroom reading instruction in the elementary grades? This question should be answered by examining the comparison of the ten studies or 43% of the sample (n=23) that used between-class ability grouping for small-group instruction. At first glance, there is a significant small mean effect of +.21 standard deviations ($p < .001$) that favors the use of between-class grouping for reading in the elementary grades. The confidence band ranges from .09 - .33, meaning there is a 95% chance that the effect falls outside of zero; children who are grouped between classrooms or grades for small-group reading instruction outperform students in non-grouped settings on tests of reading achievement.

However, there are some plausible explanations for these results. First of all, the results

are based on extremely small number of studies using a random-effects model, so any significant effects must be restricted to the populations within these studies and cannot be generalized to the wider population. Within these studies, all ten were evaluation studies of the Success For All program. There were zero studies of between-class groupings that were a non-SFA program. Since SFA is a comprehensive school reform model that features many programmatic elements, it is impossible to determine the individual contribution that between-class grouping makes to this result. This review did not uncover any SFA studies that estimate the unique contribution of between-class grouping to SFA's overall results. Consequently, we cannot speculate how the other aspects of the SFA program influence its overall success.

If we remove the SFA studies, it begs the question, why are there zero studies of between-class groupings from 1987-present? Potential reasons for the lack of non-SFA studies may be that between-class grouping, aka Joplin plans or cross-grade groupings, have become synonymous with SFA, given its long history of use. Alternatively, reviews of between-class grouping prior to 1987 suggest significant but small effects for Joplin plans (Slavin, 1987). Since this grouping format is described as both time-consuming and requiring complex coordination of services and schedules (Wallace, 1967; Slavin, 1987), it may limit strapped districts' interest in using it.

Another explanation may be that schools are using between-class groups, and it goes unreported in the literature. Anecdotally, the reviewer has worked with three schools currently or in the past in the state of Virginia that used between-class groups for reading instruction only. This suggests to the reviewer that these models are out there but are not studied except at a local level or in SFA programs. More research comparing outcomes for between- and ability-based groups is needed to understand the conditions and settings where this approach is warranted.

Overall, the second research question remains unanswered. Since there is a robust

amount of research on SFA, future research should explore the relative contributions that each of the programmatic elements provides. To the reviewer's knowledge, there are no studies of SFA that currently report the unique contributions of each program element. For future researchers, this may be a way to explore the impact of between-class grouping for reading instruction.

Research Question C

What current small-group reading instructional formats used in the general classroom have demonstrated a positive effect upon student reading achievement, e.g., differentiated instruction? The third research question shifts the synthesis' perspective from a general overview of whether small-group reading instruction affects student achievement to a detailed discussion about the type, nature and range of ability-based small-group formats used by K-5 teachers in the general classroom setting. Barr and Dreeben (1991) remind us that studies of small-group reading instruction should provide information about the nature and format of the instruction itself.

The third research question was answered during the quantitative analysis by comparing the studies according to their different instructional formats and was supported by information gleaned from the qualitative analysis of the six descriptive studies. Of all the questions, this one provided the fewest answers about the conditions of small-group instruction in the general education classroom.

Of the 29 studies included in this review, 14 reported using a specific format for small group instruction. Ten of them were SFA studies that use their own unique curricula and approach, i.e., SFA Wings. See earlier discussion under Research Question B. Although a significant small effect was found for the use of SFA *Roots and Wings*, (ES = .20, $p < .001$, CI .09 - .33), this must be interpreted to mean that SFA *Roots and Wings* had a small influence on the studies that were included in this part of the synthesis (n=10) which is unsurprising given that

these studies contributed most (10/14) of the analyzed effects. Equally surprising, was the scanty amount of research available on the various formats that teachers use for their ability-based small groups. For such a common practice, with federal policies that call for its continued use, small group reading instruction rests on a tentative empirical base.

Take the case of guided reading as an example. Although there is an abundant "how to" literature describing guided reading (Fountas & Pinnell, 1996; 1999, 2001; 2012; Schirmer & Schaffer, 2010), only one empirical field study of guided reading was located, despite its widespread acceptance and use (Ford and Opitz, 2008). Tivnan and Hemphill (2004) examined four different programs used with 590 urban first graders, among them Literacy Collaborative (LC) and Success For All. There were no significant differences found among the programs studied. Note: Literacy Collaborative is a comprehensive school reform program that uses guided reading among its programmatic features. Two teacher surveys that addressed attitudes and perceptions about guided reading were located (Ferguson & Wilson, 2009; Ford & Opitz, 2008), but there were no primary studies found that study the use of guided reading in isolation from other programmatic influences.

The PALS assessment in Virginia has provided the state's K-3 teachers with a rich pool of resources for diagnosing and planning for learning. Yet a lesser-known component of the PALS offerings is a planning framework for flexible skill-based small groups called the PALS electronic lesson plans. As with the other formats, there were no publications or studies describing their use across the state. Since a large population of Virginia teachers and teachers in other states who use the assessment and potentially, the electronic lesson planners, it would be a rich source of information if a study examining teachers' use of the electronic plans was paired with assessment results. It would also validate their use with schools in the northern parts of the state that are reluctant to use them. The PALS assessment assists the K-3 teachers who use them

to plan differentiated small-group instruction that meets children at the crossroads of the literacy stage and state standards. The PALS assessment and electronic plans adhere closely to the Virginia stage model of reading. Although aligned with research-based principles, the PALS electronic lesson planning tools would likely garner additional national attention should their use be studied and validated on a small and large scale.

Three studies identified using the Four Blocks reading program, but only one included sufficient data to calculate an effect size. Although Four Blocks has now disappeared from national prominence, another approach, Reader's Café/Daily 5 (Boushey & Moser, 2009) has ascended to the national scene. No studies were found that examined the use of Reader's Café yet, although it is newer and there may not have been sufficient time to conduct them. Reader's Café shares similarities with Four Block. Both rely on a structured routine for managing the general classroom based on claims about the use of reading-research-based principles. Both tend use a menu or "activities" based approach that is applied uniformly in the classroom. Shanahan (2014) has criticized Reader's Café as a thoughtless array of activities that pulls teachers' attention away from student outcomes. After sustained withering criticism by Moats (2007) and others, Four Blocks has all but vanished from the national scene, and the reviewer predicts that Daily 5/Reader's Café will follow a similar path.

Usually a study's outcomes tell us something important that we found in the population, but in this research synthesis, the most important finding is what we did not learn about small group reading instruction. There are not enough empirical studies validating teachers' use of small-group reading instruction. Descriptive articles and "how-to" books litter the conference floors of professional teaching organizations, but charismatic personalities and slick commercial marketing is a poor substitute for empirical evidence of "what works." Even trending and well-

established practices, such as guided reading and Daily Five/Reader's Café have small or no research support.

To summarize the findings for the research questions, there is a small but significant support for the use of ability-based small groups for reading instruction in the elementary grades. Confounding variables cloud whether the positive effects in the between-class group studies is due to grouping or other factors, and so the second research question cannot be answered. More research is needed to clarify whether between-class grouping contributes something unique beyond its role in the comprehensive reform program known as Success For All. Finally, the third question also remained a question mark due to the lack of studies that compare the formats of small group reading instruction. The primary contribution of this synthesis is to demonstrate the holes in the body of knowledge and to signal a call for more research on how small group reading instruction functions within the general K-5 classroom.

Limitations

Earlier in the chapter, alternative explanations for each of the results were presented with each research question. No study is without flaws, however, so this section addresses potential limitations of this synthesis.

Threats and protections against bias were outlined in chapter 3. The primary limitation of this study may be the narrow study inclusion criteria that forced the use of a random-effects model. Since the total number of included studies was small, this limited the usefulness of follow-up analysis and prohibited the search for mediating variables. This model limits the ability to generalize these conclusions to a wider population and suggests that the findings should be confined to this body of studies. Although no causal directions can be drawn, it is still possible to suggest relationships between the variables of interest, since the design of this research synthesis is objective and is less prone to error and data censoring than a traditional

narrative review.

Another limitation in this synthesis is the potential for missing data during coding. As research reporting standards have shifted over the last decade, many older studies did not provide sufficient details in order to calculate effect size and confidence intervals. The reviewer made every attempt to locate missing information or convert effect sizes using the data provided. If data needed for effect sizes was missing or unclear, the reviewer made notes on how effect sizes were selected. If attempts to gather effect sizes failed, but all other inclusion criteria were met, the study was set aside for qualitative review, so data would not be lost.

A third limitation is the wide variation of study designs employed by the reviewed studies. The use of effect sizes as a common metric created a standard for comparing studies with different designs. Many of the older studies used designs with serious methodological flaws. Although a rating system was used as a rough guide, determining study design quality and rigor is not an exact science. The rating guide is given to help the reader gauge the overall quality of the included studies. To mitigate error from including outcomes of poor research designs, a few precautions were used. The inclusion criteria filtered out studies that did not use experimental and quasi-experimental designs or high-quality qualitative designs that met all other criteria. The sensitivity analysis described in chapter 3 was used to insure that one study did not disproportionately skew the results.

The studies in the review employed a wide distribution of treatment variations and outcome variables for reading. Ultimately, it is impossible to tease apart whether the outcomes for any single study on grouping reflect small-group instruction or other factors not mentioned in the study, e.g., layers of intervention, coordination of whole and small group, use of professional training or special materials, etc. All of the studies in this synthesis stated in their purpose, research questions, variable descriptions, or discussions that ability-based small-group

instruction played a major role. To avoid this problem of "apples to oranges," effect sizes were used. By converting study outcomes to a single standardized measure, it was possible to make meaningful comparisons across the included studies.

Finally, the small number of included studies limits the ability to make meaningful generalization to the wider population and increases the likelihood that they are not measuring the same things. Despite this, the number of studies for elementary grades and effect sizes found in this review fell within the range of studies and effects of those reported by earlier reviews (Elbaum et al., 1999; Lou et al., 1996). The primary limitations of this study have been addressed through careful and objective procedures through each stage of the search, selection, coding, and analysis of the data.

Guidelines for Future Research

The two primary findings of this synthesis are a general positive effect for the use of small-group reading instruction during the early years of schooling, and a lack of information for the specific formats and conditions where it can be the most effective. Future studies should continue to explore the use of ability-based small groups for reading with the general population and careful study of specific formats should increase the overall effects on student achievement.

It is unclear under what conditions that small-group instruction will result in the greatest achievement gains for the most students in the general classroom. The non-significant finding for older students begs the question whether small-group instruction is simply not used as often or whether it has differential impact according to student factors (age, gender, SES, etc.), outcome measures that tap different needs at different stages of literacy, or general classroom conditions (amount time per group, group size, group movement).

With an increasingly diverse population in the U.S., the impact of small-group reading instruction on ELL populations is critical. Future studies should examine the qualities and

conditions that foster success for students placed in the lowest groups and outcomes for minorities and ELL populations must continue. The four studies that addressed this topic within the review had divergent outcomes, so more information is needed in order to have a basis for planning effective instruction and interventions.

A robust research base advocating for small-group intervention for struggling readers exists (Elbaum et al., 1999), but the results of this synthesis indicate that the same level of research does not exist for small-group instruction in the general classroom. Landmark descriptive studies of small-group instruction during the late 80's-90's (Brophy 1988; Dreeben & Barr, 1988; Flood et al., Juel, 1990; Juel & Minden-Cupp, 2000) provided nuanced portraits of the conditions and nature of instruction, but those studies are dated and lack relevance in today's classrooms. Most were based on using a basal reader with a high, middle, low group, which is not relevant to today's increasingly diverse classroom. Future studies of small group reading must take into account complex interactions of changing group membership patterns, type and dynamic quality of the instruction offered, and amount of time that is allotted to each small group session. Small-group instruction does not occur within a vacuum, so school/teacher variables must be considered, such as layers of intervention, common planning time, additional resource personnel, and administrative direction/support.

The last decade has seen advances in statistics for research that allow for analysis of multi-level growth models and the ability to test sophisticated interactions. We can create complex learning models and have the advantage of high-speed computers to assist with the analysis of effect sizes and confidence intervals. As study designs and standards for rigor continue to improve, we will design and test increasingly complicated designs that better reflect the messy chaos of classroom learning.

Current empirical studies of reading instruction in the general classroom have begun to

focus on using individualized assessment-based approaches to explore the complex interactions of planning instruction that is based on teacher, child, and instructional factors to promote reading achievement (Al Otaiba et al., 2011; Connor et al. 2011a, 2011b). There has also been a resurgence of using curriculum-based measurement to monitor and plan instruction (Deno, 1985; Deno, Fuchs, Marston, & Shin, 2001).

While these programs demonstrate gains for students in intervention, teachers in charge of large numbers of increasingly diverse students are unlikely to respond to impractical or complicated systems of learning management. Research must direct general educators to the simplest and most efficient small-group formats that are responsive to the majority of children's needs, dynamic and flexible to adjust for learning outcomes, coordinated with interventions, and easy to manage by a wide number of professionals with varying degrees of expertise and knowledge. Scripted approaches to small-group instruction that are not sensitive to children's assessed needs will not be effective if teachers lack autonomy in determining text, methods, and materials to be used (Fang, Lu, & Lamme, 2004; Tivnan & Hemphill, 2005; Wilson, Martens, & Poonam, 2005).

Direct studies of the use of between-group formats for reading instruction are warranted. Although Success For All has a proven track record for efficacy, the unique contribution of cross-grade grouping is unclear. In addition, informal reports suggest that schools that are using "home-grown" versions of SFA modeled on its programmatic principles. Direct observation and measurement of student outcomes to assess its prevalence and use are warranted since there are zero non-SFA between-group studies prior to 1987, and it is likely that the practice continues unreported in schools.

Existing studies on the efficacy of small group reading instruction formats used during K-5 instruction were meager in contrast to the popular literature that markets these formats. Well-

designed quantitative studies with comparison groups, fidelity checks, and pre-, post- designs using standardized reliable outcome measures could shed light on whether there are advantages for using one or more instructional formats during small group instruction, e.g., guided reading, Reader's Café, or PALS electronic plans over another. As Rothrock noted over 50 years ago, "Some schools are seeking for easy answers to their reading problems, usually through some magic rearrangement of the teachers or pupils; they are not likely to find the answers in this way. A well-qualified teacher may still be by far the most important factor in any grouping plan" (Rothrock, 1961, p. 235).

Implications for Theory, Policy, and Practice

Stage-based models of reading instruction suggest that children gradually acquire a vast matrix of literacy and language-based skills that appear, fade, increase, shift, or operate quietly in the eaves dependent on the child's relative standing in the reading stage and his/her assessed needs. By grouping students with similar needs together, teacher-directed small-group instruction allows a teacher to nudge a child through his/her reading stage with thoughtfully-planned instructional choices. Stage theories of reading play a critical role as guides for the nature and content of the instruction used during small groups.

State and federal policies in special and general education, namely Response to Intervention (RtI) and Common Core State Standards, have created renewed emphasis on the use of ability-groups in elementary classrooms (and beyond). The call for standards draws attention to how teachers use ability-based small-groups to plan standards-based reading instruction. One controversial aspect of the Common Core is the call for "close reading" that may influence the forms and formats of small-group ability-based instruction. Shanahan and Duffett (2013) advocate strongly for the use of challenging texts and movement away from the use of leveled texts beginning as early as second grade. This research synthesis provides support for the

continued use of ability grouping in grades K-5 and a wake-up call for educators to consider carefully what evidence exists (if any) and the conditions that support its use.

Despite the lack of studies on formats for small group instruction, this reviewer does not suggest that teachers suddenly stop using guided reading, but since small-group instruction is a common practice that is on the rise, it is critical that teachers and administrators clearly evaluate the formats that they do use in terms of the child's stage of reading and measured student outcomes to promote optimal success.

Conclusion

This research synthesis examined current practices for within- and between-class ability grouping for reading instruction in the elementary grades by reporting a set of existing empirical studies that address reading achievement outcomes and the formats of small group reading instruction. The quantitative and qualitative analysis of the effect sizes, study characteristics and themes found broad support for the use of ability-based groups for K-5 reading, in particular for within-class groups. Small but positive support was also found for the use of between-class grouping as used within the SFA program. Finally, the primary contribution of this synthesis points out the limitations of what is currently known about the efficacy and use of common formats for small group reading instruction.

So, to group or not to group? Children learn to read with the support of an adult choosing instructional texts and activities designed to lead them through the reading stages. A teacher does not have the time to meet individually with every single child, every single day. Dynamic and flexible small groups allow the teacher and children to travel the literacy landscape together-balancing the diverse demands of the classroom with the assessed behaviors and needs of individual students.

REFERENCES

Starred* citations were included in the research synthesis

Abadzi, H. (1985). Ability grouping effects on academic achievement and self-esteem: Who performs in the long run as expected. *Journal of Educational Research, 79*, 36-40.

*Adelson, J. L., & Carpenter, B. D. (2011). Grouping for achievement gains: For whom does achievement grouping increase kindergarten reading growth? *Gifted Child Quarterly, 55*(4), 265-278. doi:10.1177/0016986211417306

Allington, R. L. (1983). The reading instruction provided readers of differing reading abilities. *The Elementary School Journal, 83*, 548-559.

Allington, R. L. (2002). What I've learned about effective reading instruction from a decade of studying exemplary elementary classroom teachers. *Phi Delta Kappan, 83,* 740–747.

Allington, R. L. (2011). *What really matters in response to intervention: Research-based design* (3rd ed.). Boston: Pearson.

Al Otaiba, S., Connor, C. M., Folsom, J. S., Greulich, L., Meadows, J., & Li, Z. (2011). Assessment data-informed guidance to individualize kindergarten reading instruction. *Elementary School Journal, 111*, 535-560.

*Anderson, R. C., Wilkinson, I. A. G., & Mason, J. M. (1991). A microanalysis of the small-group, guided reading lesson: Effects of an emphasis on global story meaning. *Reading Research Quarterly, 26*(4), 417-441.

Ankrum, J., & Bean, R. (2007). Differentiated reading instruction: What and how. *Reading Horizons, 48*, 133-146.

Austin, M. C., & Morrison, C. (1963). *The first R: The Harvard report on reading in elementary schools.* New York: Macmillan.

Barr, R., & Dreeben, R. (1983). *How schools work.* Chicago: University of Chicago Press.

Barr, R., & Dreeben, R. (1991). Grouping students for reading instruction. *Handbook of Reading Research (*vol. 2*)*, pp. 885-910.

Baumann, J. F., Hoffman, J. V., Duffy-Hester, A. M., & Ro, J. M. (2000). The First R yesterday and today: US elementary reading instruction practices reported by teachers and administrators. *Reading Research Quarterly, 35*, 338-377.

Bear, D. R. (1991). Learning to fasten the seat of my union suit without looking around: The synchrony of literacy development. *Theory into Practice, 30*(3), 149-157.

Bear, D., & Barone, D. (1997). *Developing literacy: An integrated approach to assessment and instruction.* Houghton Mifflin Harcourt.

Bear, D., Invernizzi, M., Templeton, S., & Johnston, F. (2011). *Words their way: Word study for phonics, spelling and vocabulary instruction.* (5th ed.). Pearson.

Blachman, B. A., Tangel, D. M., Ball, E. W., Black, R., & McGraw, C. K. (1999). Developing phonological awareness and word recognition skills: A two-year intervention with low-income, inner-city children. *Reading and Writing, 11*(3), 239-273.

Block, J. H., & Anderson, L. W. (1975). *Mastery learning in classroom instruction.* New York: Macmillan.

Bond, G. L., & Dykstra, R. (1967). The cooperative research program in first-grade reading instruction. *Reading Research Quarterly, 2,* 5-142.

Borenstein, M., Hedges, L.V., Higgins, J.P., & Rothstein, H.R. (2011). *Comprehensive meta-analysis, version 2.2.064.* Englewood, NJ: Biostat.

Borenstein, M., Hedges, L.V., Higgins, J.P., & Rothstein, H.R. (2009). *Introduction to meta-analysis.* West Sussex, UK: John Wiley & Sons, Ltd.

Borenstein, M., Hedges, L., V., Higgins, J.P., & Rothstein, H.R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1,* 97-111.

Borman, G., & Hewes, G. (2002). The long-term effects and cost-effectiveness of Success for All. *Educational Evaluation and Policy Analysis, 24,* 243-266.

*Borman, G. D., Slavin, R. E., Cheung, A., Chamberlain, A. M., Madden, N. A., & Chambers, B. (2005). Success for all: First-year results from the national randomized field trial. *Educational Evaluation and Policy Analysis, 27*(1), 1-22.

Boushey, G., & Moser, J. (2009). *The CAFE book: Engaging all students in daily literacy assessment & instruction.* Portland, ME: Stenhouse Publishers.

Boushey, G., & Moser, J. (2014). *The daily 5: Fostering literacy independence in the elementary grades* (2nd ed.). Portland, ME: Stenhouse Publishers.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read: A causal connection. *Nature, 301,* 419-421.

Bus, A.G., van Ijzendoorn, M.H., & Mol, S. E. (2011). Meta-analysis. In N.K. Duke & M. H. Mallette. (Eds.). *Literacy research methodologies* (2nd ed.),pp. 270-300. New York: Guilford Press.

*Buttaro Jr, A., Catsambis, S., Mulkey, L., & Steelman, L. C. (2010). An organizational perspective on the origins of instructional segregation: School composition and use of within-class ability grouping in american kindergartens. *Teachers College Record, 112*(5), 1300-1336.

Calfee, R. & Brown, R. (1979). Grouping students for instruction. In S.L. Duke (Ed.), *Classroom management: Seventy-eighth yearbook of the National Society for the Study of Education Part 2*. (pp. 144-182). Chicago: University of Chicago Press.

Carson, R. M., & Thompson, J. M. (1964). The Joplin plan and traditional reading groups. *The Elementary School Journal, 65*, 38-43.

Chall, J. S. (1967). *Learning to read: The great debate* (1967). New York: McGraw Hill.

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

Chalmers, I., Hedges, L. V., & Cooper, H. (2002). A brief history of research synthesis. *Evaluation & the Health Professions, 25*(1), 12-37.

*Chambers, B., Abrami, P. C., Massue, F. M., & Morrison, S. (1998). Success for all: Evaluating an early-intervention program for children at risk of school failure. *Canadian Journal of Education, 23*(4), 357-72.

Chorzempa, B. F., & Graham, S. (2006). Primary-grade teachers' use of within-class ability grouping in reading. *Journal of Educational Psychology, 98*(3), 529-541.

Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society, 4*, 102-118.

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

Coldiron, J. R., Braddock, J. H., & McPartland, J. M. (1987). *A description of school structures and classroom practices in elementary, middle and secondary schools*. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC. CTB/McGraw-Hill.

Coleman, D., & Pimentel, S. (2012). Revised Publishers' Criteria for the Common Core State Standards in English Language Arts and Literacy, Grades 3–12. Retrieved July 20, 2014, from http://www.corestandards.org/assets/Publishers_Criteria_for_3-12.pdf.

*Condron, D. J. (2008). An early start: Skill grouping and unequal reading gains in the elementary years. *The Sociological Quarterly, 49*(2), 363-394.

Connor, C. M., Morrison, F. J., Schatschneider, C., Toste, J. R., Lundblom, E., Crowe, E. C., & Fishman, B. (2011a). Effective classroom instruction: Implications of child characteristics by reading instruction interactions on first graders' word reading achievement. *Journal of Research on Educational Effectiveness, 4*(3), 173-207. doi:10.1080/19345747.2010.510179

Connor, C. M., Morrison, F. J., Fishman, B., Giuliani, S., Luck, M., Underwood, P. S., . . . Schatschneider, C. (2011b). Testing the impact of child characteristics × instruction interactions on third graders' reading comprehension by differentiating literacy instruction. *Reading Research Quarterly, 46*(3), 189-221.

Cook, T. (2002). Randomized experiments in education: Why are they so rare? *Educational Evaluation and Policy Analysis, 24(3),* 175-199.

Cooper, H. (2009). Hypotheses and problems in research synthesis. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.). (2009), *The handbook of research synthesis* (2nd ed.). (pp. 19-37). New York: Russell Sage Foundation.

Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Los Angeles, CA: SAGE Publications.

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

*Cunningham, P., Hall, D., & McFee, M. (1991). Non-ability-grouped, multilevel instruction: A year in a first-grade classroom. *The Reading Teacher, 44,* 566-571.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52, 219–232.

Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). *School Psychology Review*, 30, 507–524.

Dreeben, R., & Barr, R. (1988a). Classroom composition and the design of instruction. *Sociology of Education, 61(3),* 129-142.

*Dreeben, R., & Barr, R. (1988b). The formation and instruction of ability groups. *American Journal of Education, 97*(1), 34-64. doi:10.1086/443912

Duffy, G. (1981). Teacher effectiveness research: Implications for the reading profession. In M.L. Kamil (Ed.) *Directions in reading: Research and instruction. 30th Yearbook of the National Reading Conference*. Washington D.C.: The National Reading Conference, Inc.

Elbaum, B., Vaughn, S., Hughes, M., & Moody, S. W. (1999). Grouping practices and reading outcomes for students with disabilities. *Exceptional Children, 65*(3), 399-415.

Fang, Z., Lu, D., & Lamme, L.L. (2004). From scripted instruction to teacher empowerment: Supporting literacy teachers to make pedagogical transitions. *Literacy, 38(1),* 58-64.

Fisher, R. A. (1932). *Statistical methods for research workers* (4th ed.). London: Oliver and Boyd.

Flesch, R. (1955). *Why Johnny can't read—and what you can do about it.* NY: Harper.

Flood, J., Lapp, D., Flood, S., & Nagel, G. (1992). Am I allowed to group? Using flexible patterns for effective instruction. *The Reading Teacher*, 608-616.

Floyd, C. (1954). Meeting children's reading needs in the middle grades: A preliminary report. *Elementary School Journal, 55*, 99-103.

Foorman, B. R., Francis, D. J., Winikates, D., Mehta, P., Schatschneider, C., & Fletcher, J. M. (1997). Early interventions for children with reading disabilities. *Scientific Studies of Reading, 1(3),* 255-276.

Foorman, B. & Torgesen, J. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. *Learning Disabilities Research & Practice, 16(4)*, 203-202.

Ford, M. P., & Opitz, M. F. (2008). A national survey of guided reading practices: What we can learn from primary teachers. *Literacy Research and Instruction, 47*(4), 309-331.

Ford, M., & Opitz, M. F., (2011). Looking back to move forward with guided reading. *Reading Horizons,* 50, 225-240.

Fountas, I. C., & Pinnell, G. S. (1996). *Guided reading: Good first teaching for all children.* Portsmouth, NH: Heinemann.

Gamoran, A., Nystrand, M., Berends, M., & LePore, P. C. (1995). An organizational analysis of the effects of ability grouping. *American Educational Research Journal, 32*, 687-715.

Gamoran, A. (1986). Instructional and institutional effects of ability grouping. *Sociology of Education, 59*(4), 185-198.

Gamoran, A. (1987). Organization, instruction, and the effects of ability grouping: Comment on Slavin's best-evidence synthesis. *Review of Educational Research, 57*(3), 341-345.

Gamoran, A. (1989). Measuring curriculum differentiation. *American Journal of Education, 97*(2), 129-143.

Gamoran, A. (1989). Rank, performance, and mobility in elementary school grouping. *The Sociological Quarterly, 30*(1), 109-123.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Gersten, R., Compton, D., Connor, C.M., Dimino, J., Santoro, L., Linan-Thompson, S., & 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 on November 1, 2014, from http://ies.ed.gov/ncee/wwc/publications/practiceguides/.

Gilrane, C. P., Russell, L. A., & Roberts, M. L. (2008). *Building a community in which everyone teaches, learns, and reads: A case study. Journal of Educational Research, 101*, 333-349.

Glass, G. V (1976). Primary, secondary and meta-analysis of research. *Educational Researcher, 10*, 3-8.

Glass, E. (2006). Meta-analysis: The quantitative synthesis of research findings. In J. L. Green, G. Camilli & P. B. Elmore. (Eds.). *Handbook of complementary methods in education research.* (pp. 427-438). Mahwah, NJ: Lawrence Erlbaum Associates.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

Goodlad, J. (1984). *A place called school: Prospects for the future*. New York: McGraw-Hill.

Green, D. & Riley, H. (1963). Interclass grouping for reading instruction. *The Journal of Experimental Education, 31*, 273-278.

*Greenlee, B. J., & Bruner, D. Y. (2001). Effects of success for all reading programs on reading achievement in title I schools. *Education, 122,* 177-188.

*Hall, D. P., & Cunningham, P. M. (1992). Reading without ability grouping: Issues in first-grade instruction. *National Reading Conference Yearbook, 41*, 235-241.

Hallinan, M. T. (1994). Tracking: From theory to practice. *Sociology of Education, 67(2),* 79-84.

*Hanselman, P., & Borman, G. D. (2013). The impacts of success for all on reading achievement in grades 3-5: Does intervening during the later elementary grades produce the same benefits as intervening early? *Educational Evaluation and Policy Analysis, 35*(2), 237-251.

Hattie, J. (2008). *Visible learning: A Synthesis of over 800 meta-analyses relating to achievement*. NY: Routledge.

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

Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Hernandez, D. (2012). *Double jeopardy: How third-grade reading skills and poverty influence high school graduation.* Baltimore, MD: Annie E. Casey Foundation.

Hiebert, E. H. (1983). An examination of ability grouping for reading instruction. *Reading Research Quarterly, 18*(2), 231-55.

Hiebert, E. H. (1987). The context of instruction and student learning: An examination of Slavin's assumptions. *Review of Educational Research, 57*(3), 337-340.

Hiebert, E. H., & Mesmer, H.A. (2013). Upping the ante of text complexity in the common core state standards: examining its potential impact on young readers. *Educational Researcher, 42*, 44-51.

Higgins, J. & Altman, D.G. (2008). Assessing risk of bias in included studies. In J. Higgins & S. Green (Eds.), *Cochrane handbook of systematic review of interventions* (pp. 187-241). West Sussex: The Cochrane Collaboration and John Wiley & Sons, Ltd.

*Hong, G. (2010). Marginal mean weighting through stratification: Adjustment for selection bias in multilevel data. *Journal of Educational & Behavioral Statistics, 35*, 499-531.

*Hong, G., Corter, C., Hong, Y., & Pelletier, J. (2012). Differential effects of literacy instruction time and homogeneous ability grouping in kindergarten classrooms: who will benefit? who will suffer? *Educational Evaluation and Policy Analysis, 34*, 69-88.

*Hong, G., & Hong, Y. (2010). Reading instruction time and homogeneous grouping in kindergarten: An application of marginal mean weighting through stratification. *Educational Evaluation and Policy Analysis, 31*(1), 54-81.

Individuals with Disabilities Education Improvement Act, U.S.C. H.R. 1350 (2004). Public Law 108-446 (20 U.S.C. 1400 et seq.)

Invernizzi, M., Meier, J., Swank, L., & Juel, C. (2004). *PALS: Phonological awareness literacy screening*. Charlottesville, VA: University of Virginia.

International Reading Association (n.d.). *International reading association: The world's leading organization of literacy professionals.* Retrieved on December 7th, 2014 from http://www.reading.org.

*Jones, C. D., & Henriksen, B. M. 1. (2013). Skills-focused small group literacy instruction in the first grade: An inquiry and insights. *Journal of Reading Education, 38(2),* 25-30.

*Juel, C. (1990). Effects of reading group assignment on reading development in 1st and 2nd-grade. *Journal of Reading Behavior, 22,* 233-254.

Juel, C. & Minden-Cupp, C. (2000). Learning to read words: Linguistic units and instructional strategies. *Reading Research Quarterly, 35*, 458-492.

Kaestle, C. (1983). *Pillars of the republic: Common schools and American society, 1780-1860.* New York: Hill & Wang.

*Kamil, M. L., & Rauscher, W. C. (1990). Effects of grouping and difficulty of materials on reading achievement. *National Reading Conference Yearbook, 39*, 121-127.

Kosanovich, M., Ladinsky, K., Nelson, L., & Torgeson, J. (2010). *Differentiated reading instruction: Small-group alternative lesson structures for all students*. Retrieved May 9, 2014, from http://www.fcrr.org/assessment/pdf/smallgroupalternativelessonstructures.pdf

Kulik, J. A (1992). *Research on ability grouping: Historical and contemporary perspectives.* Washington, DC: Office of Educational Research and Improvement.

Kulik, C. L. C., & Kulik, J. A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal, 19(3)*, 415-428.

Kulik, J. A., & Kulik, C. L. C. (1987). Effects of ability grouping on student achievement. *Equity and Excellence in Education, 23*, 22-30.

Kulik, J. A., & Kulik, C. C. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly, 36(2)*, 73-77. doi:10.1177/001698629203600204

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

*Lleras, C., & Rangel, C. (2009). Ability grouping practices in elementary school and African American/Hispanic achievement. *American Journal of Education, 115*(2), 279-304. doi:10.1086/595667

Lloyd, J. W., Forness, S. R., & Kavale, K. (1998). Some methods are more effective than others. *Intervention in School and Clinic, 33,* 195-200.

Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small-group and individual learning with technology: A meta-analysis. *Review of Educational Research, 71*, 449-521.

Lou, Y., Abrami, P. C., & Spence, J. C. (2000). Effects of within-class grouping on student achievement: An exploratory model. *Journal of Educational Research, 94*, 101-112. doi:10.1080/00220670009598748

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research, 66*, 423-458.

Loveless, T. (2013). The resurgence of ability grouping and persistence of tracking. *Part II of the Brown Center Report on American Education.* Retrieved on January 8, 201, from http://www.brookings.edu/research/reports/2013/03/18-tracking-ability-grouping-loveless.

*McCoach, D. B., O'Connell, A. A., & Levitt, H. (2006). Ability grouping across kindergarten using an early childhood longitudinal study. *Journal of Educational Research, 99,* 339-346. doi:10.3200/JOER.99.6.339-346.

*Madden, N. A., Slavin, R. E., Karweit, N. L., Dolan, L. J., & Wasik, B. A. (1993). Success for All: Longitudinal effects of a restructuring program for inner-city elementary schools. *American Educational Research Journal, 30,* 123-148.

Moats, L. (2007). Whole-language high jinks: How to tell when "scientifically-based reading instruction" isn't. *Thomas B. Fordham Institute.* Retrieved on Feb 3, 2013 from http://edex.s3-us-west-2.amazonaws.com/publication/pdfs/Moats2007_7.pdf.

Moody, S. W., Vaughn, S., & Schumm, J. S. (1997). Instructional grouping for reading: Teacher's views. *Remedial and Special Education, 18*(6), 347-356.

Morgan, E. F. & Stucker, G. B. (1960). The Joplin plan of reading versus traditional method. *Journal of Educational Psychology, 61,* 69-73.

Mundy, R. R. (2000). *The relationship of administrator-facilitated grouping for reading to student achievement and self-concept of "at risk" readers in homogeneous and heterogeneous classes* (Ed.D.). Available from ProQuest Dissertations & Theses Global. (304631012).

National Education Association. (1968). *Ability grouping research summary.* Washington, DC: Research summary 1968-3S, Research Division, National Education Association.

National Governors Association Center for Best Practices, Council of Chief State School Officers (2010). Common core state standards. National Governors Association Center for Best Practices. Council of Chief State School Officers, Washington D.C. Retrieved on February 3, 2013 from http://www.corestandards.org.

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.

Newport, J. F. (1967). The Joplin plan: The score. *The Reading Teacher, 21,*158-162.

Nobit, G.W. & Hare, R. D. (1988). *Meta-ethnography: Synthesizing qualitative studies.* (pp. 10-25, 38-47). Newbury Park, CA: SAGE Publications.

Oakes, J. (1985). *Keeping track: How schools structure inequality.* New Haven, CT: Yale University Press.

Oakes, J. (1992). Can tracking research inform practice? Technical, normative, and political considerations. *Educational Researcher, 21(4),* 12-21.

Otto, H. J. (1950). Elementary education-III: Organization and administration. In W.S. Monroe (Ed.), *Encyclopedia of Educational Research* (p.444-445). New York: Macmillan.

PALS (2011). *PALS electronic lesson plan.* Retrieved on October 1, 2013, from http://pals.virginia.edu .

Pearson, P.D. (2004). The reading wars. *Educational Policy, 18(1),* 216-252.

Pearson, P. D. (2013). Research foundations for the Common Core State Standards in English language arts. In S. Neuman and L. Gambrell (Eds.), *Reading instruction in the age of Common Core State Standards.* (pp. 237-263). Newark, DE: International Reading Association.

Polloway, E. A., Cronin, M. E., & Patton, J. R. (1986). The efficacy of group versus one-to-one instruction: A review. *Remedial and Special Education, 7*(1), 22-30.

Pope, C., Mays, N., & Popay, J. (2007). *Synthesizing qualitative and quantitative health evidence: A guide to methods.* New York: McGraw-Hill International.

Powell, W. R. (1964). The Joplin plan: An evaluation. *The Elementary School Journal, 64*, 387-392.

Publications, A. P. A., & on Journal, C. B. W. G. (2008). Reporting standards for research in psychology: why do we need them? What might they be?. *The American Psychologist, 63*(9), 839.

Quint, J. C., Balu, R., DeLaurentis, M., Rappaport, S., Smith, T. J., & Zhu, P. (2013). *The Success For All model of school reform: Early findings from the investing in innovation (I3) scale-up*. New York: MDRC.

Ramsey, W. (1962). An evaluation of a Joplin plan of grouping for reading instruction. *The Journal of Educational Research, 55*, 567-572.

*Robinson, J. P. (2008). Evidence of a differential effect of ability grouping on the reading achievement growth of language-minority Hispanics. *Educational Evaluation and Policy Analysis, 30*(2), 141-180.

Rosenholtz, S. J., & Wilson, B. (1980). The effect of classroom structure on shared perceptions of ability. *American Educational Research Journal*, *17*(1), 75-82.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 85*, 185-193.

Rosenthal, R. (1984). *Meta-analytic procedures for social research.* Newbury Park, CA: Sage Publications.

*Ross, M., Nunnery, J. A., Goldfeder, E., McDonald, A., Rachor, R., Hornbeck, M., & Fleischman, S. (2004). Using school reform models to improve reading achievement: A longitudinal study of direct instruction and success for all in an urban district. *Journal of Education for Students Placed at Risk, 9*(4), 357-388.

*Ross, S. M., & Smith, L. J. (1994). Effects of the success for all model on kindergarten through second-grade reading achievement, teachers' adjustment, and classroom-school climate at an inner-city school. *The Elementary School Journal, 95*, 121-138.

*Ross, S. M., Smith, L. J., & Casey, J. P. (1999). Bridging the gap: The effects of the Success For All program on elementary school reading achievement as a function of student ethnicity and ability level. *School Effectiveness & School Improvement, 10*(2), 129-150.

Rothrock, D. G. (1961). Heterogeneous, homogeneous, or individualized approach to reading? *Elementary English, 38,* 233-235.

Russell, D. H. (1946). Interclass grouping for reading instruction in the intermediate grades. *Journal of Educational Research, 39*, 462-470.

Scanlon, D. M., & Anderson, K. L. (2010). Using the interactive strategies approach to prevent reading difficulties in an RTI context. *Successful Approaches to RTI*, 20-65.

Scanlon, D. M., Anderson, K. L., & Sweeney, J. M. (2010). *Early intervention for reading difficulties: The interactive strategies approach.* New York: Guilford.

Scanlon, D. M., Gelzheiser, L. M., Vellutino, F. R., Schatschneider, C., & Sweeney, J. M. (2008). Reducing the incidence of early reading difficulties: Professional development for classroom teachers versus direct interventions for children. *Learning and Individual Differences, 18*(3), 346-359.

Scarcelli, S. M., & Morgan, R. F. (1999). The efficacy of using a direct reading instruction approach in literature based classrooms. *Reading Improvement, 36*(4), 172-179.

Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., Shavelson, R. J. (2007). *Estimating causal effects using experimental and observational designs* (Report from the Governing Board of the American Educational Research Association Grants Program). Washington, DC: American Educational Research Association.

Shanahan, T. (2001). Research synthesis: Making sense of the accumulation of knowledge in reading. In M. Kamil, P. Mosenthal, P.D. Pearson (Eds.). *Handbook of reading research, (vol. 3).* (pp. 209-226). Routledge.

Shanahan, T. (2014). *How to organize daily literacy instruction -- Part II.* Retrieved on December 7, 2014 from http://www.shanahanonliteracy.com/search/label/Daily%20Five

Shanahan, T., & Duffett, A. (2013). *Common core in the schools: A first look at reading assignments.* Retrieved on October 3, 2014, from http://www.edexcellence.net/publications/common-core-in-the-schools.

Simpson, R. J. S., & Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *The British Medical Journal, 2,* 1243-1246.

Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research, 57*(3), 293-336. Retrieved from http://www.jstor.org/stable/1170460

Slavin, R. E. (1995). Detracking and its detractors. *Phi Delta Kappan, 77,* 220-221.

Slavin, R. (2010). Effective reading programs for Title 1 schools. Retrieved from http://gradelevelreading.net/wp-content/uploads/2010/10/Slavin-Research-Review-of-Reading-Programs-for-Title-I-Schools.pdf .

*Slavin, R. E., Madden, N. A., Karweit, N. L., Livermon, B. J., & Dolan, L. (1990). Success for all: First-year outcomes of a comprehensive plan for reforming urban education. *American Educational Research Journal, 27,* 255–278.

Smith, N. B. (1965). *American reading instruction.* Newark, DE. International Reading Association.

*Smith, L. J., Ross, S. M., & Casey, J. (1996). Multi-site comparison of the effects of success for all on reading achievement. *Journal of Literacy Research, 28*(3), 329-53.

Spear-Swerling, L., & Sternberg, R. J. (1996). *Off-track: When poor readers become "learning disabled".* Boulder, CO: Westview Press.

Spectrum K12 School Solutions. (2009). *Response to intervention (RTI) adoption survey 2009.* Retrieved on September 15, 2014, from http://www.spectrumk12.com .

Stock, W.A. (1994). Systematic coding for research synthesis. In H. Cooper & L.V. Hedges (Eds.). *The Handbook of Research Synthesis* (pp.41-56). New York: Russell Sage Foundation.

Stradling, B., & Saunders, L. (1993). Differentiation in practice: Responding to the needs of all pupils. *Educational Research, 35,* 127-137.

Success For All Foundation© (2012). Success For All Foundation. Retrieved on November 1, 2014, from http://www.successforall.org .

Success For All Foundation©. (2012). *Cost effectiveness of SFA* [Article]. Retrieved November 1, 2014, from http://www.successforall.org/cost-effective/.

*Tach, L. M., & Farkas, G. (2006). Learning-related behaviors, cognitive skills, and ability grouping when schooling begins. *Social Science Research, 35*(4), 1048-1079.

Taylor, B. M., Hanson, B. E., Justice-Swanson, K., & Watts, S. M. (1997). Helping struggling readers: Linking small-group intervention with cross-age tutoring. *The Reading Teacher*, 196-209.

Taylor, B. M., Short, R. M., Shearer, B. A., & Frye, B. J. (1995). First grade teachers provide reading intervention in the classroom. In R. L. Allington and S. A. Walmsley, (Eds.). *No Quick Fix: Rethinking Literacy Programs in America's Elementary Schools* (pp.159-176). New York: Teachers College Press.

Taylor, B. M., Pearson, P. D., Clark, K., & Sharon Walpole. (2000). Effective schools and accomplished teachers: Lessons about primary-grade reading instruction in low-income schools. *The Elementary School Journal, 101*, 121-165.

Tieso, C. L. (2003). Ability grouping is not just tracking anymore. *Roeper Review, 26*(1), 29-36.

Tieso, C. (2005). The effects of grouping practices and curricular adjustments on achievement. *Journal for the Education of the Gifted, 29*(1), 60-89.

*Tivnan, T. & Hemphill, L. (2005). Comparing four literacy reform models in high-poverty schools: Patterns of first-grade achievement. *The Elementary School Journal, 105*, 419-441.

Tomlinson, C. (1999). Mapping a route toward differentiated instruction. *Educational Leadership, 57*(1), 12-16.

Tomlinson, C. (1999). *The differentiated classroom: Responding to the needs of all learners.* Alexandria, VA: Association for Supervision and Curriculum Development.

Tomlinson, C. (2001). *How to differentiate instruction in mixed-ability classrooms,* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Tomlinson, C. (2008). *Differentiated instruction.* Retrieved August 10, 2014, from http://www.caroltomlinson.com.

Tomlinson, C., Brighton, C., Hertberg, H., Callahan, C., Moon, T.R., Brimijoin, K., Conover, L.A., & Reynolds, T. (2003). Differentiating instruction in response to student readiness, interest, and learning profile in academically diverse classrooms: A review of the literature. *Journal for the Education of the Gifted, 27(2/3),* 119-145.

Tyack, D. B. (1974). *The one best system: A history of American urban education.* Harvard University Press.

Tyner, B. (2009). Small-group reading instruction: A differentiated teaching model for beginning and struggling readers. (2nd ed.). Newark, DE: International Reading Association.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (n.d.) Retrieved from http://nces.ed.gov/nationsreportcard/.

U.S. Department of Education, National Center for Education Statistics. (2000). *ECLS-K Base Year Restricted-Use Child File* (CD-ROM and user's manual). (NCES 2000-097). Washington, DC: Author.

Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). Tutorial: How many studies do you need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics, 35*(2), 215-247.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., & Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88*(4), 601-638.

Vellutino, F. R., Scanlon, D. M., Zhang, H., & Schatschneider, C. (2008). Using response to kindergarten and first grade intervention to identify children at-risk for long-term reading difficulties. *Reading and Writing, 21*(4), 437-480.

Vygotsky, L. S. (1987). Thinking and speech. In L. S. Vygotsky, Collected works (vol. 1, pp. 39–285) (R. Rieber & A. Carton, Eds; N. Minick, Trans.). New York: Plenum. (Original works published in 1934, 1960).

Walpole, S., & McKenna, M. C. (2007). *Differentiated reading instruction: Strategies for the primary grades.* New York: Guilford Press.

Washburne, C.W. (1953). Adjusting the program to the child. *Educational Leadership, 11(3),* 138-147.

What Works Clearinghouse (2007). *Beginning reading topic report.* Retrieved from http://ies.ed.gov/ncee/wwc/documentsum.aspx?sid=27

White, H.D. (1994). Scientific communication and literature retrieval. In Cooper & L.V. Hedges (Eds.). *The Handbook of Research Synthesis* (pp.41-56). New York: Russell Sage Foundation.

Wilson, B.J. & Schmits, D.W. (1978). What's new in ability grouping. *Phi Delta Kappan, 59,* 535-536.

Wilson, D. B. (n.d.). *Practical meta-analysis effect size calculator.* Retrieved on August 28, 2014, from *http://www.campbellcollaboration.org/resources/effect_size_input.php.*

Wood, D. J., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychiatry and Psychology,* 17(2), 89-100.

Wilson, P., Martens, P., & Poonam, A. (2005). Accountability for reading and readers: What the numbers don't tell. *The Reading Teacher, 58(7)*, 622-631.

Appendix A

Coding Manual

Study Feature	Type of Feature	Description
1. Study ID	Study Identification	A unique number assigned to each study for identification and retrieval purposes taken from Refworks file.
2. Author	Study Identification	First author's last name & 2nd author (if applicable)
3. Pub_Year	Study Identification	Year of publication. Enter ? if you cannot find this information.
4. Abstract	Study Identification	1-3 sentences summarizing key points of study
5. Research question	Study Identification	Concise summary of research question/s
6. Inclusion Criteria	Inclusion /Exclusion Criteria	Initial Screening Criteria: Does the study meet **all** of these conditions: 1. Does the study use an experimental or quasi-experimental or qualitative research design? 2. Does it address small-group reading instruction in grades K-5 for the regular classroom? 3. Was it published between 1987-present and written in English? **If YES, then answer the remaining questions** **If not sure, then answer the remaining questions & talk to the reviewer** Inclusion Criteria: 1) Did the study use an experimental or quasi-experimental research design? 2) Does it focus on small-group instruction for reading in K-5 in the **general** classroom or Tier 1 setting? *Note: It does not focus on intervention or special education. It does not include Tier 2 or 3 instruction. It may be conducted by a classroom teacher, specialist or instructional assistant as long as it occurs within the general classroom during the language arts block. It does not focus on content area reading instruction.*

		3) Does the study use within-class or between-class ability-based reading groups? 4)Does the study address reading instruction during the elementary grades? Elementary grades is defined as Kindergarten- fifth grade. 5)Does the study describe the instructional formats used during small group instruction? 6)Does the study report a reading achievement measure as one or more of the outcomes, i.e., global reading achievement, word identification or comprehension? *Note: Answer no if the study measures reading attitudes/engagement only.* 7) Does the study report or allow for the calculation of effect sizes?
7. Eligible for Systematic Review?	Meets Exclusion Criteria	Yes= meets all, include in review QY= qualified yes –meets all but criteria #5 If unresolved after discussion with reviewer, include in review. No= exclude from analysis and list the reason/s
8. Type of Publication	Study Identification	What type of publication was this from? 1= Journal Article 2= Book or chapter 3= Dissertation or MA Thesis 4= Government report (local, state, or federal) 5= Conference report or presentation 6=Private report or online report 7=Specify other:_____ 8= Not Sure or Unknown
9. Peer Review	Study Identification	Did the study undergo the peer review process in order to be selected for publication? 1 = Yes 2= No 3= Not sure or Unknown
10. Method	Study Quality	If the paper describes a study performed by the authors, identify the type of research design used by the authors for the study: 1= Experimental (included a *randomly sampled* and *randomly assigned* treatment and control group; dependent & independent variables were defined; took steps to insure validity and reliability and eliminate bias; may include between subjects, within subjects or factorial

		designs) 2= Quasi-experimental (treatment/control groups but random assignment &/or random sampling *was not used*, e.g., classrooms without assigning teachers or students randomly to the treatment conditions, convenience sample; may include pre-post testing of matched or repeated measures) 3= Qualitative study (observational, case study, ethnography, survey, descriptive) 4= Correlational studies (naturalistic observation, survey, reports relationships as correlations between independent var's and outcomes) 5= Narrative Literature Review 6= Systematic Review or Meta-analysis 7= Not specified / unknown
11. Overall Quality of Study Design_1	Study Quality	1= presence of experimental and control group (or matched design or repeated measure) 0 = absent or no control group used ? = Not reported or not sure
12. Overall Quality of Study Design_2	Study Quality	2= Used random sampling and selection 1= compared matched or non- random samples, 0= Not reported
13. Overall Quality of Study Design_3	Study Quality	1= Reported comparability of treatment & control group at pre- and post 0 = Did not report comparability of treatment & comparison group at pretesting and post-testing
14. Overall Quality of Study Design_4	Study Quality	1= reported a fidelity check for treatment 0 = no fidelity check for treatment is reported
15. Overall Quality of Study Design_5	Study Quality	2= published with peer review 1= published without review process (online ERIC document, conference brief/report) 0 = unpublished manuscript /MA thesis/dissertation
16. Level of School	Setting	1 = Elementary school (includes K-5) 2= Middle School/High School (gr. 6-12) 3= Not reported
17. School/Class Location	Setting	0= Unknown 1= Urban 2= Rural 3= Suburban 4= Includes 2 or more settings, e.g., Urban & Rural
18. Group_Inst_Location	Subjects	Where did students receive most of their small-group reading instruction in the study?

		1= General education classroom 2= Special education or Title I classroom 3= Gifted classroom 4=Combination or multi-age classrooms 5= Unknown /Not reported
19. Unit Under Study	Subjects	What unit/s/ does the study focus on or report? 1=Student 2=Classroom 3=Teacher 4=School 5= Two or more of the above. Specify: 6= Unknown or not reported
20. Student Gender	Subjects	1= Male 2=Female 3= Both M& F 0= Not reported
21. Ethnicity	Subjects	1= African American 2=Caucasian/White 3= Hispanic 4= Asian 5= Native American 5=2 or more ethnicities reported. 6= Not reported
22. Socioeconomic Status (SES)	Subjects	Does the study focus on students who are… 1= Low SES only 2=Average SES only 3=High SES only 4=Two or more levels included 0=Not reported or unknown
23. Grade Levels	Subjects	What grade/s were the subjects in the study? 0 = Kindergarten only 1= First grade only 2= Second grade only 3= Third grade only 4= Fourth grade only 5= Fifth grade only 6= Range of grades. Specify: K-3 or 3-5 7= unknown or not reported
24. Total Sample Size	Method	What was the total sample size for the study before attrition or missing data? N=
25. Treatment n	Method	What was the total sample size for the treatment group? E =
26. Control n	Method	What was the total sample size for the control or comparison group? C =

27. Ability Group Type	Treatment	What type of ability grouping was used for the treatment group? 1= Within-class grouping 2= Between-class grouping 3= Not reported or unknown
28. Avg_Ability_Grp	Treatment	What is the average # of ability groups used by the classroom teacher? 0= zero groups 1= 1 group 2= 2 groups 3 = 3 groups 4= 4 or more groups NR = not reported
29. Ability_Group_Size	Treatment	How many students were assigned to each small group? Report the average # . 1= <5 per group 2= 5-8 per group 3= 8-10 per group 4= >10 per group 5= Not reported
30. Grp_Assignmt_Criteria	Treatment	What criteria was used for group membership assignment? 1= Instructional reading level 2= Reading strategies 3= Specific reading skills: (specify) 4= Other: specify _____ 5=Not reported
31. Small Group Format	Treatment	What was the primary organizational format used during small-group reading instruction? 1= Guided Reading (Fountas & Pinnell) 2= Daily 5/**Reader's Café** 3=PALS Electronic Plans 4= Differentiated Instruction 5= Basal/core program 6= Interactive Strategy Approach 7= Four Blocks 6= Other: specify & describe 7= Not Reported or Unknown
32. Type of Sm_Grp Instructional Format	Treatment	How would you categorize the primary instructional format for the treatment group? 1= Text-centered supported reading format 2= Skills-based lesson 3= Other: specify 4= Not Reported
33. Grp_Membership Change	Treatment	Did the small group membership remain the same throughout the treatment implementation?

		1= Yes 2= No 3= Not reported or Unknown
34. Prof`lDev	Treatment	Did the teachers in the experimental group receive more or different professional development training than controls? 1= Yes 2= No 3= Not reported
35. Curr_Materials	Treatment	Did the experimental group receive more or different curricular materials from the control group? 1= Yes 2= No 3= Not reported
36. Duration of Study	Treatment	What is the duration of the study's treatment? Length in weeks (NR = not reported. Estimate 1 school year = 36 weeks)
37. Sm_Gp_Length	Treatment	What is the average number of minutes allotted for each small group to meet, i.e., how long did teachers meet with each group? Report average # of minutes or NR= not reported/unknown)
38. Sm_Gp_Days	Treatment	How many days/week did the small groups meet for instruction? Report the average # of days per week: _____ (NR= not reported)
39. LABlock_Total_Time	Treatment	What is the total length of time allotted for daily reading block or language arts instruction? 0 = none 1 =1 hour or less 2 = 1-2 hours 3= 2 or more hours NR = not reported
40. IndepVar	Treatment	What is the main variable that is deliberately changed by the study?
41. IndepVar2	Treatment	What were the additional independent variables in the study (if any)?
42. Control_Var	Treatment	Control Variables = any other variable that may affect the outcome that needed to be kept the same or monitored throughout the study.
43. Depend_Var1	Outcome	Dependent Variables = What is the primary or first outcome variable called that is measured by the study?

44. DV1_Type	Outcome	Dependent Variable1_Type of Measure is… 1 = Standardized normed instrument 2= Researcher-made for study 3= Teacher-made informal test
45. DV1_RdgArea	Outcome	Dependent Variable1_Reading Area that is measured… 1= Phonological awareness or phonemic awareness skill 2= Phonics or alphabet (letter name, sound, production) 3= High frequency or sight words 4= Word identification skills (may include nonsense word lists) 5= Passage reading/fluency (instructional reading level determined by IRI, timed reading of passages for accuracy and wpm) 6= Vocabulary 7= Comprehension 8= Global Reading (combined two or more into a single assessment for overall measure of reading 9= Other: specify: 10= Not reported or unknown
46. DV1_TestName		Dependent Var.1_Name of Standardized test = What is the name of the standardized outcome used in the study (if applicable)
47. DV2_Type 48. DV3_Type 49. DV4_Type	Outcome	Describe the type of measure used for additional Dept var's in the study. 1 = Standardized normed instrument 2= Researcher-made for study 3= Teacher-made informal test
50. DV2_RdgArea 51. DV3_RdgArea 52. DV4_RdgArea	DV_Outcome Measure (area of reading tapped):	**Use for any additional Dept var's listed in the study.** Dependent Variable2-4_Reading Area that is measured… 1= Phonological awareness or phonemic awareness skill 2= Phonics or alphabet (letter name, sound, production) 3= High frequency or sight words 4= Word identification skills (may include nonsense word lists) 5= Passage reading/fluency (instructional reading level determined by IRI, timed reading of passages for accuracy and wpm) 6= Vocabulary 7= Comprehension

		8= Global Reading (combined two or more into a single assessment for overall measure of reading 9= Other: specify: 10= Not reported or unknown
53. DV2_TestName 54. DV3_TestName 55. DV4_TestName	Outcome	**Use for any additional Dept var's in the study.** Dependent Var.2-4_Name of Standardized test = What is the name of the standardized outcome used in the study (if applicable)
56. Treatment Mean	Outcomes	Report the study's means for the treatment group for each DV. Specify which variable by DV1= , DV2= , DV3= etc.
57. Treatment Standard Deviation	Outcomes	Record the standard deviation for the treatment group for each DV. Specify which variable by DV1= , DV2= , DV3= etc.
58. Control Mean	Outcomes	Record the control group's means on each DV (if available). Specify which variable by DV1=, DV2= , DV3= etc.
59. Control Standard Deviation	Outcomes	Record the control group's standard deviation for each DV (if available). Specify which variable by DV1= , DV2= , DV3= etc.
60. Effect Size Reported by Study	Effect Size	Record any effect sizes reported by the study.
61. Effect size Calculated by Coder	Effect Size	Calculate using CMA software or online calculator with reviewer. Use means, standard deviation, t-, p, or F-values or correlations/standard error.
62. Attrition of Subjects	Study Quality	Does the study report and address how they handled subjects who dropped out of the study? Yes= briefly describe Not Reported
63. Missing Data	Study Quality	Does the study report and address how they handled any missing data? Yes = briefly describe Not Reported
64. Comments/ Questions	Reviewer Notes & Questions	Any notes or questions about the study that were not captured by the codes.

Appendix B

Characteristics of Included Studies

Study Design (Type) Description	Type of Grouping	Small Group Format	Grade/s	Reading Outcome	Sample Size	Mean Effect Size	Study Design	Study Quality/Rigor (0-12)
Adelson, J. & Carpenter, B., 2011 (journal) Used ECLS-K data to analyze gifted vs non-gifted classes for low, mid, high. Found sig. positive for mid/high compared & sig.for low groups.	Within	Not reported	Kind.	Global Measure	9340 students in 1690 classes	Low= -.06 Mid=+1.34 High= +1.47	Q-Exp	9
+Anderson, Wilkinson & Mason, 1991 (journal) Micro-analysis of small group reading instruction using basal series for code-emphasis vs. word-emphasis	Within	Guided Reading*	3rd	Comprehension	149 students in 6 classes	NR	Qual.	6
Borman, Slavin, Cheung, et al. 2005 (journal) National randomized study of SFA vs 'regular" instruction in 35 schools	Between	SFA Roots & Wings	K-3	Alphabet Phonics Word Id/Wd Attack Comprehension	35 sites with SFA = 1085 Control= 1023	Mean ES for ABC = .16 Comp = .16 Wd Id/ Attack = +.27	Exp.	12

Study Design (Type) Description	Type of Grouping	Small Group Format	Grade/s	Reading Outcome	Sample Size	Mean Effect Size	Study Design	Study Quality/Rigor (0-8)
Buttaro & Catsambis et al. 2010 (journal) Used ECLS-K data to compare group achievement with overall school/organization focus	Within	NR	K	Global Reading Time spent in Groups	2022 teachers 546 schools	Mean ES +.39	Q-Exp	9
Chambers, Abrami, Massue, & Morrison, 1998 (journal) Independent evaluation of SFA in 4 urban Canadian schools (1 SFA, 3 controls)	Between	SFA Roots & Wings	K-6	Wd Attack/Word Id Reading Comprehension	425 students from 4 schools	Overall Mean +.17	Q-Exp	9
Condron, 2008 (journal) Uses ECLS-K to examine grouping for 1st and 3rd graders (low, mid, high groups)	Within	NR	1-3rd	Global Reading	N= 3654 (ability) and 11,588 (non-grouped)	Mean ES = -2.08	Q-Exp	9
+Cunningham, Hall & Defee, 1998 (journal) Rationale and study of 4 Block in a 1st grade reading program	Within	Four Block	1st	Passage Reading/IRI	24	NR	Q-Exp/ qual.-- mixed method	4

Study Design (Type) Description	Type of Grouping	Small Group Format	Grade/s	Reading Outcome	Sample Size	Mean Effect Size	Study Design	Study Quality/Rigor (0-8)
+Dreeben & Barr, 1988 (journal) Descriptive study of 1st grade grouping for reading focused on organizational factors and classroom composition.	Within	Basal	1st	Letters & Phonics Word Identification Global Reading	3 districts, 7 schools, 13 classes, 50 groups, >300 1st graders.	NR	Qual. - descriptive	10
Greenlee & Bruner, 2001 (journal) Independent evaluation that compares SFA to regular Title I small group instruction. Finds negative impact against SFA.	Between	SFA	1-5th	Global Reading	N=34 schools SFA trtmt = 11 Title 1 control = 23	Mean ES = -.877 (against SFA)	Q-Exp	6
+Hall & Cunningham, 1992 (journal) 4 Block in 2 first grade classrooms	Within	4 Block	1st	Passage Reading/IRI	6 classes of 22-26 students each	NR	Qual.	4

Study Design (Type) Description	Type of Grouping	Small Group Format	Grade/s	Reading Outcome	Sample Size	Mean Effect Size	Study Design	Study Quality/Rigor (0-8)
Hanselman & Borman, 2013 (journal) Cluster-randomized trial in gr. 3-5 of SFA – found n.s. differences	Between	SFA Roots & Wings	3-5	Global Reading	N= 35 schools, 2240 and 2172 in 35 schools	Mean ES = .02	Q-Exp	11
Hong & Hong, 2010 (journal) Uses ECLS-K data to examine ability group placement in kindergarten (high vs low/mid)	Within	NR	K	Global Reading	N=8668	ES = +.24	Q-Exp	11
Hong, Corter, Hong & Pelletier, 2012 (journal) Use ECLS-K data to investigate kindergarten group placement and intensity/length of small group time. Found significant effects for low and middle. Small but n.sig. for high.	Within	NR	K	Global Reading	N=8668	All Groups (vs none) = +.12 Low = .27 Middle = .16	Q-Exp	11
Hong_ 2010 (journal) Demonstrates use of marginal mean weighting and uses ECLS-K to within-class ability groups in kindergarten. Finds positive results for grouping.	Within	NR	K	Global Reading	N=1863 across intensity groups	+.16	Q-Exp	11

Study Design (Type) Description	Type of Grouping	Small Group Format	Grade/s	Reading Outcome	Sample Size	Mean Effect Size	Study Design	Study Quality/Rigor (0-8)
+Jones & Henricksen, 2013 (journal) Descriptive action-research study of flexible skills-based small group reading effects on at-risk and at/above learners	Within	Flexible-Skills-based groups	1st	Nonsense Word Oral Reading Fluency	25 students in 1 classroom	Insufficient to calculate	Q-Exp	4
+Juel, 1990 (journal) Describes effects & type of reading group instruction in 1st-2nd grade	Within	Basal	1st-2nd	Global Reading Alphabet/ Phonics Phon. Awareness	80	NR	Qual.	9
Kamil & Rauscher, 1990 (journal) Compares ability v. whole group groups & level of reading materials	Within	Basal	3-5th	Vocabulary Comprehension	869 3-5th grade students	-1.79	Q-Exp	6
Lleras & Rangel, 2009 (journal) ECLS-K study of how ability grouping affects Hispanics & AAmericans achievement	Within	NR	1st, 3rd	Global Reading	AA = 750 Hispanic = 886	MEAN ES AAmer = -.17 Hisp = -.14	Q-Exp	11

Study Design (Type) Description	Type of Grouping	Small Group Format	Grade/s	Reading Outcome	Sample Size	Mean Effect Size	Study Design	Study Quality/Rigor (0-8)
Madden, Slavin, Karweit, Dolan, & Wasik, 1993 (journal) Presents 3 years of data on SFA in 5 Baltimore urban schools matched with controls.	Between	SFA Roots & Wings	1-3	Wd Attack/Wd Id. Oral Reading	1st =309 2nd = 276 3rd = 257	Mean ES +.69	Q-Exp	10
McCoach, O'Connell & Leavitt, 2006 (journal) ECLS-K study of ability groups in kindergarten. Examined groups vs. no grouping. Found positive effects for Kind. Grouping.	Within	NR	K	Global Reading	10,191	ES = +.25 (small)	Q-Exp	10
Robinson, 2008 (journal) Used ECLS-K data to compare ability grouping for Hispanics Found positive effects for grouping.	Within	NR	K-3	Global Reading	@1600	Overall .172	Q-Exp	10
Ross, Nunnery, Goldfeder, et al., 2004 (journal) 3rd party evaluation comparing SFA to Direct Instruction in 9 elem. schools. Non.sig results for both SFA & DI compared to controls	Between (SFA)	SFA Roots & Wings	2-4	Global Reading Not-significant for SFA or DI compared to control "traditional" schools	9 schools N= 3939 (cell n's were not given)	Overall Gr. 4 = -.13 Gr 2 = +.12	Mixed design	10

Study Design (Type) Description	Type of Grouping	Small Group Format	Grade/s	Reading Outcome	Sample Size	Mean Effect Size	Study Design	Study Quality/Rigor (0-8)
Smith, Ross, & Casey _1996 (journal) Independent multi-site comparison of SFA in 4 cities Found positive results in 3 sites especially for bottom 25%ile of students.	Between	SFA	K-2	Global Reading	K-2 students in 4 schools	+.50	Q-Exp	11
Tach & Farkas, 2006 (journal) Uses ECLS-K to examine group placement and reading achievement in Kindergarten	Within	NR	K-1	Global Reading	756	Mean ES 1.275	Q-Exp	11
Tivnan & Hemphill, 2005 (journal) Compared outcomes of 4 different reading programs that use small groups (Building Essential Literacy (BEL); Developing Literacy First; Literacy Collaborative; and Success For All)	Between & Within	SFA (between groups) Guided Reading (within groups)	1st	Global Reading	590 urban 1st graders	SFA (.28) Guided Rdg (.54)	Q-Exp	7

Note: N= 29 All reported ES represent the overall mean for the study or independent samples analyzed within the study. Sign (+/-) indicates direction of effect for the type of grouping. Study quality and rigor were assessed using 0-12 point scale, with scores closer to 0 representing weak designs and scores closer to 12 representing stronger designs. Criteria are described under Quality Measures of Included Studies.
Key: NR = not reported and insufficient data to calculate ES; SFA = Success For All; T= Treatment, C = Control
Type of Grouping: Within, Between, No Grouping
Formats: Basal, Four Block, Guided Reading, PALS Electronic Lesson Plans, Reader's Café, Differentiated Instruction
List of included studies are marked with a * in the References.

Study Design (Type) Description	Type of Grouping	Small Group Format	Grade/s	Reading Outcome	Sample Size	Mean Effect Size	Study Design	Study Quality/Rigor (0-8)
Ross & Smith, 1994 (journal) Evaluates SFA on 1 school after 1 year of implementation. Sign for kind but n.s. for 1st/2nd.	Between	SFA Roots & Wings	K-2	Word attack/ Word identification Global Reading Measure Comprehension	220 (109 SFA and 111 matched controls)	Overall +.09	Q-Exp	11
Ross, Smith, & Casey, 1999 Evaluates impact of SFA on minorities and lowest 25%ile in 2 midwestern schools. N.S. between SFA and control school but did find sign. diff between minorities that favored SFA students over controls.	Between	SFA Roots & Wings	3-5	Global Reading	SFA= 157 and Control = 99	Overall +.13	Q-Exp	11
Scarcelli & Morgan, 1999 (journal) Compare 4 Block vs. whole language classes in 1st	Within	4 Block	1st	Global Reading	N= 55 in 4 classes 4Block=25 WL = 30	Overall .23	Q-Exp	7
Slavin, Madden, Karweit et al. 1990 (journal) Reporting 1st year outcomes of SFA in one K-3 against matched control	Between	SFA (Roots & Wings)	K-3	Global Reading	N=456 Matched pairs in Kind = 60 1st = 59 2nd = 65 3rd = 44	Overall K-3 = .50	Q-Exp	11

APPENDIX C

List of Excluded Studies

	Excluded Studies N=214	**Exclusion Reason**
1	Aarnouts, C., Van Leeuwe, J., Voeten, M., & Oud, H. (2001). Development of decoding, reading comprehension, vocabulary and spelling during the elementary school years. *Reading and Writing: An Interdisciplinary Journal,* **14:** 61–89, 2001.	Focus is not on small group instruction
2	Al Otaiba, S., Connor, C. M., Folsom, J. S., Greulich, L., Meadows, J., & Li, Z. (2011). Assessment data-informed guidance to individualize kindergarten reading instruction. Elementary School Journal, 111(4), 535-560.	Focus is on individualized instruction
3	Allington, R. L. (1983). The reading instruction provided readers of differing reading abilities. The Elementary School Journal, 83(5), 548-559.	Published before 1987
4	Alperin, S. L. (2004). Observations of reading ability groups within a second-grade classroom: A case study (Ed.D.). Available from ProQuest Dissertations & Theses Global. (305047761).	Unpublished dissertation – based on 1 classroom only
5	Alpert, B., & Bechar, S. (2008). School organizational efforts in search for alternatives to ability grouping. Teaching and Teacher Education, 24(6), 1599-1612. doi:10.1016/j.tate.2008.02.023	Used secondary students
6	Amendum, S. J., Li, Y., Hall, L. A., Fitzgerald, J., Creamer, K. H., Head-Reeves, D., & Hollingsworth, H. L. (2009). Which reading lesson instruction characteristics matter for early reading achievement? Reading Psychology, 30(2), 119-147. doi:10.1080/02702710802275173	Focus is not on small groups
7	Ansalone, G. (2010). Tracking: Educational differentiation or defective strategy. Educational Research Quarterly, 34(2), 3-17.	Used middle grade students and focus on social studies
8	Ansalone, G., & Ming, C. (2006). Programming students for academic success: The PLS an alternative to traditional tracking. (undetermined). Educational Research Quarterly, 29(3), 3-10.	Not a study – reviews tracking
9	Assel, M. A., Landry, S. H., & Swank, P. R. (2007). An evaluation of curriculum, setting, and mentoring on the performance of children enrolled in pre-kindergarten. Reading & Writing, 20(5), 463-494. doi:10.1007/s11145-006-9039-5	Uses Pre-K students and grouping is not the focus
10	Avalos, M. A., Plasencia, A., Chavez, C., & Rascón, J. (2008). Modified guided reading: Gateway to english as a second language and literacy learning. The Reading Teacher, 61(4), 318-329.	Uses middle school students
11	Baker, L., & Wigfield, A. (1999). Dimensions of children's motivation for reading and their relations to reading activity and reading achievement. Reading Research Quarterly, 34(4), 452-477. doi:10.1598/RRQ.34.4.4	Focus on multi-age classroom assignment for principals
12	Barr, R., & Dreeben, R. (1991). Grouping students for reading instruction. Handbook of Reading Research, 2, 885-910.	Literature review
13	Bear, D. R., & Barone, D. (1989). Using children's spellings to group for word study and directed reading in the primary classroom. Reading Psychology,	Focus is not on small group

	10(3), 275-292. doi:10.1080/0270271890100305	
14	Bennett, K. P. (1987). a study of reading ability grouping and its consequences for urban appalachian first grade students (Educat.D.). Available from ProQuest Dissertations & Theses Global. (303463497).	Unpublished dissertation –method flaws
15	Biemiller, A. (1993). Lake wobegon revisited: On diversity and education. Educational Researcher, 22(9), 7-12.	Not a study
16	Bond, G. L., & Dykstra, R. (1997). The cooperative research program in first-grade reading... (cover story). Reading Research Quarterly, 32(4), 348.	Descriptive – focus is not on small group
17	Borland, J. H., Horton, D., & Subotnik, R. F. (2002). Ability grouping and acceleration of gifted students: Articles from the roeper review. Roeper Review, 24(3), 100-101. doi:10.1080/02783190209554138	Literature review
18	Boushey, G., & Moser, J. (2006). The daily 5 : Fostering literacy independence in the elementary grades. Portland, Me.: Stenhouse Publishers.	Not a study – focused on independent work
19	Bracey, G. W. (1986). Ability grouping and student achievement in elementary schools. The Phi Delta Kappan, 68(1), 76-77.	Not a study
20	Bracey, G. W. (1993). Against ability grouping... again. The Phi Delta Kappan, 74(7), 573-574.	Not a study
21	Braddock,Jomills Henry,,II, Slavin, R. E., & Center for Research on Effective Schooling for,Disadvantaged Students. (1992). Why ability grouping must end: Achieving excellence and equity in american education	Literature review
22	Brady, S., Fowler, A., Stone, B., & Winbury, N. (1994). Training phonological awareness: A study with inner-city kindergarten children. Annals of Dyslexia, 44, 26-59.	Intervention
23	Brown, J., & Morris, D. (2005). Meeting the needs of low spellers in a second-grade classroom. Reading & Writing Quarterly: Overcoming Learning Difficulties, 21(2), 165-184. doi:10.1080/10573560590915969	Intervention – focus is not small group
24	Burns, M. K. (2007). Reading at the instructional level with children identified as learning disabled: Potential implications for response-to-intervention. School Psychology Quarterly, 22(3), 297-313. doi:10.1037/1045-3830.22.3.297	Intervention
25	Burns, M. K., & Ysseldyke, J. E. (2005). Comparison of existing responsiveness-to-intervention models to identify and answer implementation questions. *California School Psychologist, 10*, 9-20.	Focus is not small group
26	Burns, R. B., & Mason, D. A. (1995). Organizational constraints on the formation of elementary school classes. American Journal of Education, 103(2), 185-212.	Focus on classroom composition not small groups
27	Burns, R. B., & Mason, D. A. (2002). Class composition and student achievement in elementary schools. American Educational Research Journal, 39(1), 207-233.	Focus on multi-age classroom assignment for principals
28	Bursuck, B., & Blanks, B. (2010). Evidence-based early reading practices within a response to intervention system. Psychology in the Schools, 47(5), 421-431.	Focus on intervention and RTI
29	Boushey, G., & Moser, J. (2009). *The CAFE book : Engaging all students in daily literacy assessment & instruction*. Portland, Me.: Stenhouse Publishers ;Markham, Ont.	Not a study – describes a series of activities
30	Byrd, R. C. (2006). A comparison of multiage classrooms (Ed.D.). Available	Unpublished

	from ProQuest Dissertations & Theses Global. (304913085).	dissertation on ability-based classrooms
31	Caldwell, J. S., & Ford, M. P. (2002). Where have all the bluebirds gone? how to soar with flexible grouping	Not a study
32	Callahan, M., Griffo, V. B., & Pearson, P. D. (2009). Teacher knowledge and teaching reading. *College Reading Association Yearbook,* (30), 37-62.	Not a study
33	Canady, C. E., & Canady, R. L. (2012). Catching readers up before they fail. Educational Leadership, 69(9), 1-4.	Not a study
34	Cantrell, S. C. (1998). Effective teaching and literacy learning: A look inside primary classrooms. Reading Teacher, 52(4), 370.	Not a study
35	Carol McDonald Connor, , ,Lara J. Jakobsons, Elizabeth C. Crowe, a., & Meadows, J. G. . (2009). Instruction, student engagement, and reading skill growth in reading first classrooms. The Elementary School Journal, 109(3), 221-250.	Focus on individualized teacher-managed instruction
36	CARPENTER, K., GEHSMANN, K., SMITH, R., BEAR, D., & TEMPLETON, S. (2009). Learning together: Putting word study instruction into practice. California Reader, 42(3), 4-18	Focus is not on small group
37	Carson, R. M., & Thompson, J. M. (1964). The joplin plan and traditional reading groups. The Elementary School Journal, 65(1), 38-43.	Published before 1987
38	Catsambis, S., & Buttaro, A.,Jr. (2012). Revisiting "kindergarten as academic boot camp": A nationwide study of ability grouping and psycho-social development. *Social Psychology of Education, 15(4),* 483-515. doi:10.1007/s11218-012-9196-0	Focus on attitudes
39	Catsambis, S., Sophia.Catsambis@qc.cuny.edu, Mulkey, L. M. 2., Buttaro, A., Steelman, L. C., & Koch, P. R. (2012). Examining gender differences in ability group placement at the onset of schooling: The role of skills, behaviors, and teacher evaluations. *Journal of Educational Research, 105*, 8-20. doi:10.1080/00220671.2010.514779	Focus on behavior and atttitudes.
40	Chambers, B., Slavin, R. E., Madden, N. A., Cheung, A., Gifford, R., & Success for, A. F. (2005). Enhancing success for all for hispanic students: Effects on beginning reading achievement. Success for All Foundation.	Unpublished report
41	Chard, D. J., & Kameenui, E. J. (2000). Struggling first-grade readers: The frequency and progress of their reading. Journal of Special Education, 34(1), 28.	Focus on intervention
42	Cheng, R. W., Lam, S., & Chan, J. C. (2008). When high achievers and low achievers work in the same group: The roles of group heterogeneity and processes in project-based learning. British Journal of Educational Psychology, 78, 205-221. doi:10.1348/000709907X218160	Focus is not on reading
43	Chorzempa, B. F., & Graham, S. (2006). Primary-grade teachers' use of within-class ability grouping in reading. Journal of Educational Psychology, 98(3), 529-541. doi:10.1037/0022-0663.98.3.529	Survey of teacher attitudes – no reading outcomes
44	Cirino, P. T., Pollard-Durodola, S. D., Foorman, B. R., Carlson, C. D., & Francis, D. J. (2007). Teacher characteristics, classroom instruction, and student literacy and language outcomes in bilingual kindergartners. The Elementary School Journal, 107(4), 341-364.	Focus on teacher quality and characteristics for bilingual instruction – not small groups

45	Condron, D. J. (2005). *Stratification, skill grouping, and learning to read in first grade* (Ph.D.). Available from ProQuest Dissertations & Theses Global. (305399562).	Unpublished dissertation - results are published later and included in synthesis
46	Condron, D. J. (2007). Stratification and educational sorting: Explaining ascriptive inequalities in early childhood reading group placement. *Social Problems, 54*(1), 139-160.	Focus is on race and gender to predict small group placement – no focus on instruction
47	Connor, C. M., Morrison, F. J., Fishman, B., Giuliani, S., Luck, M., Underwood, P. S., . . . Schatschneider, C. (2011). Testing the impact of child characteristics × instruction interactions on third graders' reading comprehension by differentiating literacy instruction. Reading Research Quarterly, 46(3), 189-221.	Focus on individualized instruction
48	Connor, C. M., Morrison, F. J., Schatschneider, C., Toste, J. R., Lundblom, E., Crowe, E. C., & Fishman, B. (2011). Effective classroom instruction: Implications of child characteristics by reading instruction interactions on first graders' word reading achievement. Journal of Research on Educational Effectiveness, 4(3), 173-207. doi:10.1080/19345747.2010.510179	Focus on individualized instruction
49	Cunningham, P. M., Hall, D. P., & Defee, M. (1991). Non-ability-grouped, multilevel instruction: A year in a first-grade classroom. Reading Teacher, 44, 566-571.	Not a study – describes Four Block
50	Curby, T. W., Stuhlman, M., Grimm, K., Mashburn, A., Chomat-Mooney, L., Jason Downer, . . . Pianta, R. C. (2011). Within-day variability in the quality of classroom interactions during third and fifth grade. The Elementary School Journal, 112(1), 16-37.	Grouping is one variable but not the focus of study
51	Denton, C. A., Kethley, C., Nimon, K., Kurz, T. B., Mathes, P. G., Shih, M., & Swanson, E. A. (2010). Effectiveness of a supplemental early reading intervention scaled up in multiple schools. *Exceptional Children, 76*(4), 394-416.	Intervention study
52	Dreeben, R., & Barr, R. (1988). Classroom composition and the design of instruction. *Sociology of Education, 61*(3), 129-142. Doi:10.2307/2112622	Descriptive of class composition. Not a study
53	Duffy, G. G. (1981). Teacher effectiveness research: Implications for the reading profession.	Literature review
54	Eastman, V. L. (2010). Small group instruction: Reading instruction utilizing learning style preferences and the reading achievement of first grade students (Ph.D.). Available from ProQuest Dissertations & Theses Global. (305186345).	Unpublished dissertation focused on learning styles not small groups
55	Eder, D. (1983). Ability grouping and students' academic self-concepts: A case study. The Elementary School Journal, 84(2), 149-161. doi:10.1086/461352	Focus is on attitude
56	Elbaum, B. E., Schumm, J. S., & Vaughn, S. (1997). Urban middle-elementary students' perceptions of grouping formats for reading instruction. Elementary School Journal, 97(5), 475-500. doi:10.1086/461877	Focus is on attitude
57	Elbaum, B., Moody, S. W., & Schumm, J. S. (1999). Mixed-ability grouping for reading: What students think. Learning Disabilities Research & Practice,	Focus is on attitude

	14(1), 61-66. doi:10.1207/sldrp1401_6	
58	Elbaum, B., Vaughn, S., Hughes, M., & Moody, S. W. (1999). Grouping practices and reading outcomes for students with disabilities. Exceptional Children, 65(3), 399-415.	Literature review
59	Fawson, P. C., & Reutzel, D. R. (2000). But I only have a basal: Implementing guided reading in the early grades. The Reading Teacher, 54(1), 84-97.	About guided reading but not a study
60	Ferrell, L. J. (2000). Literacy development within multiage and single grade age cohorts: The impact of organizational structure (Ed.D.). Available from ProQuest Dissertations & Theses Global. (304614065).	Unpublished dissertation focus on multi-age classrooms
61	Fiedler, E. D., Lange, R. E., & Winebrenner, S. (2002). In search of reality: Unraveling the myths about tracking, ability grouping, and the gifted. Roeper Review, 24(3), 108-111. doi:10.1080/02783190209554142	Not a study
62	Ferguson, J., & Wilson, J. (2009). Guided reading: It's for primary teachers. *College Reading Association Yearbook,* (30), 293-306.	Qualitative survey of teacher reports on guided reading – no achievement scores
63	Fisher, C. W., & Hiebert, E. H. (1990). Characteristics of tasks in two approaches to literacy instruction. The Elementary School Journal, 91(1), 3-18.	Focus is on literacy skills not groups
64	Flood, J., & Lapp, D. (1992). Am I allowed to group? using flexible patterns for effective instruction. Reading Teacher, 45(8), 608.	Not a study
65	Foorman, B. R., & Torgesen, J. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. Learning Disabilities Research & Practice (Wiley-Blackwell), 16(4), 203.	Literature review
66	Ford, M. P., & Opitz, M. F. (2008). A national survey of guided reading practices: What we can learn from primary teachers. *Literacy Research and Instruction, 47(4)*, 309-331.	Not a study –survey of teacher perceptions
67	Ford, M. P., & Opitz, M. F. (2011). Looking back to move forward with guided reading. *Reading Horizons, 50*, 225-240.	History of guided reading
68	Fountas, I. C., & Pinnell, G. S. (2012). Guided reading: The romance and the reality. Reading Teacher, 66(4), 268-284. doi:10.1002/TRTR.01123	Not a study – describes guided reading
69	Fountas, I. C., & Pinnell, G. S. (c1996.). Guided reading : Good first teaching for all children. Portsmouth, NH : Heinemann,.	Not a study – describes guided reading
70	Fountas, I. C., & Pinnell, G. S. (c1999.). Matching books to readers : Using leveled books in guided reading, K-3 / irene C. fountas and gay su pinnell. Portsmouth, NH : Heinemann,.	Not a study – describes guided reading
71	Fountas, I. C., & Pinnell, G. S. (c2001.). Guiding readers and writers, grades 3-6 : Teaching comprehension, genre, and content literacy. Portsmouth, N.H. : Heinemann,.	Not a study – describes guided reading
72	Gamoran, A. (1986). Instructional and institutional effects of ability grouping. Sociology of Education, 59(4), 185-198. doi:10.2307/2112346	Published prior to 1987
73	Gamoran, A. (1987). Organization, instruction, and the effects of ability grouping: Comment on slavin's "best-evidence synthesis". Review of Educational Research, 57(3), 341-345.	Not a study
74	Gamoran, A. (1989). Measuring curriculum differentiation. American Journal	Not a study

	of Education, 97(2), 129-143. doi:10.1086/443918	
75	Gamoran, A. (1989). Rank, performance, and mobility in elementary school grouping. The Sociological Quarterly, 30(1), 109-123.	Focus is not on small group
76	Gamoran, A., Nystrand, M., Berends, M., & LePore, P. C. (1995). An organizational analysis of the effects of ability grouping. American Educational Research Journal, 32(4), 687-715.	Uses middle grades
77	Garzella, M. F. (1991). Using an expert system to diagnose weaknesses and prescribe remedial reading strategies among elementary learning-disabled students (Ed.D.). Available from ProQuest Dissertations & Theses Global. (304008865).	Focus on intervention
78	Gelzheiser, L. M., & Meyers, J. (1991). Reading instruction by classroom, remedial, and resource room teachers. Journal of Special Education, 24, 512-526. doi:10.1177/002246699102400409	Focus is on teacher /spsecialist coordination
79	Gelzheiser, L. M., Scanlon, D., Vellutino, F., Hallgren-Flynn, L., & Schatschneider, C. (2011). Effects of the interactive strategies Approach—Extended: A responsive and comprehensive intervention for intermediate-grade struggling readers. The Elementary School Journal, 112(2), 280-306.	Focus is 1:1 tutoring
80	Gentry, M., & Owen, S. V. (1999). An investigation of the effects of total school flexible cluster grouping on identification, achievement, and classroom practices. Gifted Child Quarterly, 43(4), 224-243. doi:10.1177/001698629904300402	Focus is on high school achievement
81	Gillies, R. M., & Ashman, A. F. (1998). Behavior and interactions of children in cooperative groups in lower and middle elementary grades. Journal of Educational Psychology, 90(4), 746-757. doi:10.1037/0022-0663.90.4.746	Focus is on cooperative learning (heterogeneous groups)
82	Goldenberg, C. N. (1989). Parents' effects on academic grouping for reading: Three case studies. American Educational Research Journal, 26(3), 329-352.	Focus is on parent involvement
83	Good, T. L. (1997). Do students learn more in heterogeneous or homogeneous groups? Document Resume Ed 268 075 Sp 027 044, , 27.	Unpublished report
84	Green, J. L., Camilli, G., Elmore, P. B., & American Educational Research Association. (2006). *Handbook of complementary methods in education research*. Mahwah, N.J.: Lawrence Erlbaum Associates ;Washington, D.C.	Not a study
85	Guastello, E. F., & Lenz, C. (2005). Student accountability: Guided reading kidstations. Reading Teacher, 59(2), 144-156. doi:10.1598/RT.59.2.4	Not a study
86	Gutiérrez, R., & Slavin, R. E. (1992). Achievement effects of the nongraded elementary school: A best evidence synthesis. Review of Educational Research, 62(4), 333-376.	Literature review
87	Hallam, S., & Parsons, S. (2013). The incidence and make up of ability grouped sets in the UK primary school. Research Papers in Education, 28(4), 393-420. doi:10.1080/02671522.2012.729079	Focus on class composition
88	Haller, E. J. (1985). Pupil race and elementary school ability grouping: Are teachers biased against black children? American Educational Research Journal, 22(4), 465-483. doi:10.2307/1163135	Focus on race and attitudes
89	Hallinan, M. T., Bottoms, E., Aaron M. Pallas, & Aaron M. Palla. (2003). Ability grouping and student learning. Brookings Papers on Education Policy, (6), 95-140.	Not a study

90	Harp, B. (1989). What do we know now about ability grouping. *Reading Teacher, 42*(6), 430-431.	Not a study
91	Harrison, A. E. (1989). Ability grouping: Practices and perceptions of elementary school teachers (Ed.D.). Available from ProQuest Dissertations & Theses Global. (303706667).	Unpublished dissertation
92	Haskins, R., Walden, T., & Ramey, C. T. (1983). Teacher and student behavior in high- and low-ability groups. Journal of Educational Psychology, 75(6), 865-876. doi:10.1037/0022-0663.75.6.865	Published prior to 1987
93	Hensel, L. S. (2009). The convergence of the belief systems and instructional practices of special education teachers during reading instruction (Ph.D.). Available from ProQuest Dissertations & Theses Global. (305140567).	Unpublished dissertation
94	Hiebert, E. H. (1983). An examination of ability grouping for reading instruction. Reading Research Quarterly,	Published prior to 1987
95	Hiebert, E. H. (1987). The context of instruction and student learning: An examination of slavin's assumptions. Review of Educational Research, 57(3), 337-340.	Not a study
96	Hong, G. (2009). Marginal mean weighting through stratification: Adjustment for selection bias in multilevel data. Journal of Educational & Behavioral Statistics, 35(5), 499-531. doi:10.3102/1076998609359785	Focus is on statistical method not reading – same results published elsewhere
97	Hooper, S., Ward, T. J., Hannafin, M. J., & Clark, H. T. (1989). The effects of aptitude composition on achievement during small-group learning. Journal of Computer-Based Instruction, 16(3), 102-109.	Focus on attitudes
98	Hutchison, D. (2003). The effect of group-level influences on pupils' progress in reading. British Educational Research Journal, 29(1), 25.	Focus on classroom assigment
99	Iaquinta, A. (2006). Guided reading: A research-based response to the challenges of early reading instruction. Early Childhood Education Journal, 33(6), 413-418. doi:10.1007/s10643-006-0074-2	Not a study- describes guided reading
100	Institute of, E. S. (2007). Success for all. what works clearinghouse intervention report. What Works Clearinghouse.	Online review
101	Invernizzi, M., Abouzeid, M., & Gill, J. T. (1994). Using students' invented spellings as a guide for spelling instruction that emphasizes word study. The Elementary School Journal, 95(2), 155-167.	Focus is not small group
102	Ireson, J., & Hallam, S. (1999). Raising standards: Is ability grouping the answer? Oxford Review of Education, 25(3), 343-358.	Not a study
103	Irwin, L. S. (1998). A reading specialist's story of reflection and collaboration: An endeavor to support peers as they practice flexible grouping to meet students' literacy learning needs (Ed.D.). Available from ProQuest Dissertations & Theses Global. (304478201).	Unpublished dissertation
104	Jacobson, J. E. (1990). A national study of elementary principals' perceptions of reading education issues (Ed.D.). Available from ProQuest Dissertations & Theses Global. (303801411).	Focus on administrator attitudes
105	Jacobson, J., Reutzel, D., & Hollingsworth, P. (1992). Reading-instruction - perceptions of elementary-school principals. Journal of Educational Research, 85(6), 370-380.	Focus on administrator attitudes
106	Jecks, D. (2011). Flexible skills grouping of elementary students for reading	Unpublished

	instruction: A meta-analytic review (Ed.D.). Available from ProQuest Dissertations & Theses Global. (916923169).	dissertation
107	Johns, R. B. (1982). The friday plan. The Phi Delta Kappan, 63(9), p. 630.	Published prior to 1987
108	Jongsma, E. (1985). Research views: Grouping for instruction. The Reading Teacher, 38(9), 918-920.	Published prior to 1987
109	Juel, C., & Minden-Cupp, C. (2000). Learning to read words: Linguistic units and instructional strategies. Reading Research Quarterly, 35(4), 458-492.	Qualitative descriptive study of 1st grade instruction
110	Kinnucan-Welsch, K., Rosemary, C. A., & Grogan, P. R. (2006). Accountability by design in literacy professional development. *Reading Teacher, 59*(5), 426-435. doi:10.1598/RT.59.5.2	Not a study
111	Klein, A. M. (2005). The connection between social skills and academic achievement among latino and caucasian elementary school children (Ph.D.). Available from ProQuest Dissertations & Theses Global. (305003627).	Unpublished dissertation
112	Kletzien, S. B. (1996). Reading programs in nationally recognized elementary schools. Reading Research & Instruction, 35, 260-274.	Focus is not on small group instruction
113	Kulik, J. A., & Kulik, C. C. (1992). Meta-analytic findings on grouping programs. Gifted Child Quarterly, 36(2), 73-77. doi:10.1177/001698629203600204	Meta-analysis
114	Kulikand, J. A., & Kulik, C. C. (1987). Effects of ability grouping on student achievement. Equity & Excellence in Education, 23(1-2), 22-30. doi:10.1080/1066568870230105	Lit. review
115	Lanning, L. (2002). Managing and monitoring the rest of the class during guided reading: Three literacy educators share their stories. New England Reading Association Journal, 38(3), 20-24.	Description of 3 classrooms – not an empirical study
116	Lofton, K. J. C. (2013). An examination of skill groups, traditional groups, and attendance patterns in third-grade classrooms: Their effects on student achievement based on discovery assessment test C scores (Ed.D.). Available from ProQuest Dissertations & Theses Global. (1453479825).	Unpublished dissertation – math study
117	Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. Review of Educational Research, 66(4), 423-458.	Literature review
118	Lyle, S. (1999). An investigation of pupil perceptions of mixed-ability grouping to enhance literacy in children aged 9-10. Educational Studies, 25(3), 283-296. doi:10.1080/03055699997800	Focus on class composition
119	Macintyre, H., & Ireson, J. (2002). Within-class ability grouping: Placement of pupils in groups and self-concept. British Educational Research Journal, 28(2), 249-263. doi:10.1080/01411920120122176	Focus on class composition
120	Macqueen, S. (2012). Academic outcomes from between-class achievement grouping: The australian primary context. Australian Educational Researcher, 39(1), 59-73. doi:10.1007/s13384-011-0047-3	Focus on middle grades
121	Madden, N. A., Slavin, R. E., Karweit, N. L., & Livermon, B. J. (1989). Restructuring the urban elementary-school. Educational Leadership, 46(5), 14-18.	Not a study – review article
122	Madden, N. A., Slavin, R. E., Karweit, N. L., Dolan, L., & Wasik, B. A.	Not a study

	(1991). Success for all. Phi Delta Kappan, 72(8), 593-599.	
123	Marsh, H. W. (1984). Self-concept, social comparison, and ability grouping: A reply to kulik and kulik. American Educational Research Journal, 21(4), 799-806. doi:10.2307/1163002	Not a study
124	Mason, D. A., & Doepner, R. W. (1998). Principals' views of combination classes. Journal of Educational Research, 91, 160-172. doi:10.1080/00220679809597537	Survey of administrators – focus on multiage classes
125	Mathes, P. G., & Torgesen, J. K. (1998). All children can learn to read: Critical care for the prevention of reading failure. Peabody Journal of Education, 73(3/4, Literacy Education in the 21st Century), 317-340.	Literature review
126	Matthews, M. S., Ritchotte, J. A., & McBee, M. T. (2013). Effects of schoolwide cluster grouping and within-class ability grouping on elementary school students' academic achievement growth. High Ability Studies, 24(2), 81-97. doi:10.1080/13598139.2013.846251	Focus on between class placement (homogeneous classroom)
127	Matthews, M. W., Monsaas, J. A., & Penick, J. M. (1997). A comparative study of the literacy development of at-risk children in graded versus nongraded classrooms	Unpublished manuscript
128	McCoach, D. E. (2003). Does grouping matter? A cross-classified random effects model of children's reading growth during the first two years of school (Ph.D.). Available from ProQuest Dissertations & Theses Global. (305344939).	Unpublished dissertation – later published report is included in study
129	McCurdy, M., Daly, E., & Gortmaker, V. (2007). Use of brief instructional trials to identify small group reading strategies: A two experiment study. Journal of Behavioral Education, 16(1), 7-26. doi:10.1007/s10864-006-9021-y	Focus on individualized instruction
130	McGee, L. M., & Morrow, L. M. (2005). Teaching literacy in kindergarten Guilford Publications.	Not a study
131	Melser, N. A. (1999). Gifted students and cooperative learning: A study of grouping strategies. Roeper Review: A Journal on Gifted Education, 21(4), 315. doi:10.1080/02783199909553983	Focus on heterogeneous groups
132	Miller, B. (1991). The effects of ability grouping and classroom climate on student achievement in selected texas schools (Ed.D.). Available from ProQuest Dissertations & Theses Global. (303952812).	Unpublished dissertation
133	Mills, C. J., & Durden, W. G. (1992). Cooperative learning and ability grouping - an issue of choice. Gifted Child Quarterly, 36(1), 11-16. doi:10.1177/001698629203600103	Focus on cooperative learning
134	Moats, L. (2007). Whole-language high jinks: How to tell when" scientifically-based reading instruction" isn't. Thomas B.Fordham Institute,	Not a study
135	Moody, S. W., Vaughn, S., & Schumm, J. S. (1997). Instructional grouping for reading - teacher's views. Remedial and Special Education, 18(6), 347-356.	Focus on attitudes
136	Morris, D., Blanton, L., Blanton, W. E., & Perney, J. (1995). Spelling instruction and achievement in six classrooms. The Elementary School Journal, 96(2), 145-162.	Focus is skills not small group
137	Mosenthal, J., Lipson, M., Torncello, S., Russ, B., & Mekkelsen, J. (2004). Contexts and practices of six schools successful in obtaining reading achievement. The Elementary School Journal, 104(5), 343-367.	Focus is not on small groups
138	Mouzaki, A., Foorman, B. R., & Santi, K. (2005). Preventing reading	Online report - focus

	problems: The application of an assessment-driven, classroom-based intervention in texas schools. (pp. 21-49) Learning Disabilities Worldwide.	is intervention
139	Mundy, R. R. (2000). The relationship of administrator-facilitated grouping for reading to student achievement and self-concept of "at risk" readers in homogeneous and heterogeneous classes (Ed.D.). Available from ProQuest Dissertations & Theses Global. (304631012).	Unpublished dissertation – intervention study
140	Neihart, M. (2007). The socioaffective impact of acceleration and ability grouping: Recommendations for best practice. Gifted Child Quarterly, 51(4), 330-341. doi:10.1177/0016986207306319	Not a study
141	Oakes, J. (1992). Can tracking research inform practice? technical, normative, and political considerations. Educational Researcher, 21(4), 12-21.	Not a study
142	Onchwari, G., & Keengwe, J. (2010). Teacher mentoring and early literacy learning: A case study of a mentor-coach initiative. *Early Childhood Education Journal, 37*(4), 311-317.	Focus is not small group
143	Palardy, G. J., & Rumberger, R. W. (2008). Teacher effectiveness in first grade: The importance of background qualifications, attitudes, and instructional practices for student learning. Educational Evaluation and Policy Analysis, 30(2), 111-140.	Focus is not on small groups
144	Pallas, A. M., Entwisle, D. R., Alexander, K. L., & Stluka, M. F. (1994). Ability-group effects - instructional, social, or institutional. Sociology of Education, 67(1), 27-46. doi:10.2307/2112748	Focus is on school organization
145	Petrilli, M. (2011). All together now? educating high and low achievers in the same classroom. Education Next, 11(1), 48-55.	Not a study
146	Piasta, S. B., & Wagner, R. K. (2010). Developing early literacy skills: A meta-analysis of alphabet learning and instruction. Reading Research Quarterly, 45(1), 8-38.	Literature review
147	Plesa, S. E. (2004). A qualitative study concerning instructional grouping of students in reading (Ed.D.). Available from ProQuest Dissertations & Theses Global. (305034498).	Unpublished dissertation
148	Plesa, S. E. (2004). A qualitative study concerning instructional grouping of students in reading (Ed.D.). Available from ProQuest Dissertations & Theses Global. (305034498).	Unpublished dissertation
149	Poole, D. (2008). Interactional differentiation in the mixed-ability group: A situated view of two struggling readers. Reading Research Quarterly, 43(3), 228-250. doi:10.1598/RRQ.43.3.2	Focus on cooperative mixed groups
150	Porche, M. V., Pallante, ,Daniel H., & Snow, C. E. (2012). Professional development for reading AchievementResults from the collaborative language and literacy instruction project (CLLIP). The Elementary School Journal, 112(4), 649-671.	Not a study of small groups
151	Powell, W. R. (1964). The joplin plan: An evaluation. The Elementary School Journal, 64(7), 387-392.	Published priot to 1987
152	Puzio, K., Colby, G., & Society for Research on,Educational Effectiveness. (2010). The effects of within class grouping on reading achievement: A meta-analytic synthesis. ().Society for Research on Educational Effectiveness.	Unpublished literature review
153	Rambo-Hernandez, K. E. (2011). How much do schools matter? using summer growth patterns to assess the impact of schools on high achieving and gifted	Unpublished dissertation – not

	students (Ph.D.). Available from ProQuest Dissertations & Theses Global. (915016932).	focused on small groups
154	Rashotte, C. A., MacPhee, K., & Torgesen, J. K. (2001). The effectiveness of a group reading instruction program with poor readers in multiple grades. Learning Disability Quarterly, 24(2), 119-134.	Intervention study
155	Rasinski, T. V., & Others, A. (1992). *The effects of fluency development instruction on reading for urban second grade students.* ().	Focus is not on small group
156	Raudenbush, S. W., & Bryk, A. S. (1988). Methodological advances in analyzing the effects of schools and classrooms on student learning. Review of Research in Education, 15, 423-475.	Focus is not small groups
157	Rees, D. I., Argys, L. M., & Brewer, D. J. (1996). Tracking in the united states: Descriptive statistics from NELS. Economics of Education Review, 15(1), 83-89. doi:10.1016/0272-7757(95)00025-9	Literature review
158	REEVES, S., BISHOP, J., & FILCE, H. G. (2010). Response to intervention (rtl) and tier systems: Questions remain as educators make challenging decisions. *Delta Kappa Gamma Bulletin, 76*(4), 30-35.	Intervention – not a study
159	Reis, S. M., McCoach, D. B., Coyne, M., Schreiber, F. J., Eckert, R. D., & Gubbins, E. J. (2007). Using planned enrichment strategies with direct instruction to improve reading fluency, comprehension, and attitude toward reading: An evidence-based study. Elementary School Journal, 108(1), 3-23.	Intervention/enrichment study
160	Reis, S. M., McCoach, D. B., Coyne, M., Schreiber, F. J., Eckert, R. D., & Gubbins, E. J. (2007). Using planned enrichment strategies with direct instruction to improve reading fluency, comprehension, and attitude toward reading: An evidence-based study. Elementary School Journal, 108(1), 3-23.	Enrichment/ intervention
161	Reis, S. M., McCoach, D. B., Little, C. A., Muller, L. M., & Kaniskan, R. B. (2011). The effects of differentiated instruction and enrichment pedagogy on reading achievement in five elementary schools. American Educational Research Journal, 48(2), 462-501.	Enrichment/ intervention
162	Reutzel, D. R. (2011). Organizing literacy classrooms for effective instruction. Reading Teacher, 65(2), 96-109. doi:10.1002/TRTR.01013	Not a study
163	Robbins, P. (2006). A quasi-experimental retrospective study of third-grade small group reading interventions (Ed.D.). Available from ProQuest Dissertations & Theses Global. (304913680).	Unpublished dissertation - intervention
164	Rogers, S., Danielson, K., & Russell, J. F. (2000). Collaborating to promote effective elementary practices across seven school districts. Peabody Journal of Education, 75(3, Collaboration: Across Campus, across Town, and with K-12 Schools), 133-144.	Focus is not on grouping or small groups
165	Rosenholtz, S. J., & Wilson, B. (1980). The effect of classroom structure on shared perceptions of ability. American Educational Research Journal, 17(1), 75-82.	Published prior to 1987
166	Ross, S. M., & Others, A. (1994). Using "success for all" to restructure elementary schools: A tale of four cities.	Online report
167	Ross, S. M., & Smith, L. J. (1994). Effects of the success for all model on kindergarten through second-grade reading achievement, teachers' adjustment, and classroom-school climate at an inner-city school. The Elementary School Journal, 95(2), 121-138.	Focus on attitude and climate

168	Ross, S. M., Smith, L. J., Lohr, L., & McNelis, M. (1994). Math and reading instruction in tracked first-grade classes. *The Elementary School Journal, 95*(2), 105-119.	Mixed Method – observations & surveys. No small group observed.
169	Rowan, B., & Miracle, A. W. (1983). Systems of ability grouping and the stratification of achievement in elementary schools. Sociology of Education, 56(3), 133-144. doi:10.2307/2112382	Published prior to 1987
170	Rowan, B., Miracle, A. W., Jr., & Far West Lab for Educational Research,and Development. (1982). Systems of ability grouping and the stratification of achievement in elementary schools.	Published prior to 1987
171	Sacks, C. H., & Mergendoller, J. R. (1997). The relationship between teachers' theoretical orientation toward reading and student outcomes in kindergarten children with different initial reading abilities. American Educational Research Journal, 34(4), 721-739.	Focus is not small group instruction
172	Schirmer, B. R., & Schaffer, L. (2010). Guided reading approach. Teaching Exceptional Children, 42(5), 52-58.	Not a study
173	Schumm, J. S., Moody, S. W., & Vaughn, S. (2000). Grouping for reading instruction: Does one size fit all? Journal of Learning Disabilities, 33(5), 477-488. doi:10.1177/002221940003300508	Focus on intervention
174	Schwartz, R. M. (2005). Decisions, decisions: Responding to primary students during guided reading. The Reading Teacher, 58(5), 436-443.	Not a study
175	Shake, M. C., & Allington, R. L. (1985). Where do teachers' questions come from? The Reading Teacher, 38(4), 432-438.	Published prior to 1987
176	Simmons, D. C., & Kuykendall, K. (2000). Implementation of a schoolwide reading improvement model: `No one ever told us it would be this.. Learning Disabilities Research & Practice (Lawrence Erlbaum), 15(2), 92.	Focus is not on grouping
177	Slavin, R. (2002). Mounting evidence supports the achievement effects of success for all. Phi Delta Kappan, 83(6), 469-480.	Literature review
178	Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. Review of Educational Research, 57(3), 293-336.	Literature review
179	Slavin, R. E. (1987). Grouping for instruction in the elementary school. Educational Psychologist, 22(2), 109.	Review article
180	Slavin, R. E. (1995). Detracking and its detractors: Flawed evidence, flawed values. Phi Delta Kappan, 77(3), 220-221.	Review article
181	Slavin, R. E. (1998). Reading by nine: What will it take? Peabody Journal of Education, 73(3/4, Literacy Education in the 21st Century), 68-80.	Review article
182	Slavin, R. E., & Madden, N. A. (2001). Reducing the gap: Success for all and the achievement of african-american and latino students. ().	Online report
183	Slavin, R. E., & Madden, N. A. (2006). Reducing the gap: Success for all and the achievement of african american students. Journal of Negro Education, 75(3), 389-400.	Review article
184	Slavin, R. E., & Others, A. (1996). Success for all: A summary of research. Journal of Education for Students Placed at Risk, 1(1), 41-76.	Review article
185	Slavin, R. E., Center for Research on Effective Schooling for,Disadvantaged Students, & Others, A. (1992). Success for All. policy implications. report no.	Online policy report

	35..	
186	Slavin, R. E., Madden, N. A., & Dolan, L. J. (1994). 'Whenever and wherever we choose: The replication of 'success for all'. Phi Delta Kappan, 75, 639.	Review article
187	Sørensen, A. B., & Hallinan, M. T. (1986). Effects of ability grouping on growth in academic achievement. American Educational Research Journal, 23(4), 519-542. doi:10.2307/1163088	Published prior to 1987
188	Spear-Swerling, L., & Cheesman, E. (2012). Teachers' knowledge base for implementing response-to-intervention models in reading. *Reading and Writing: An Interdisciplinary Journal, 25*(7), 1691-1723.	Focus is intervention
189	Sterbinsky, A. D. (2003). Reliability of the school observation measure using generalization theory and the frequency of classroom practices used in schools (Ph.D.). Available from ProQuest Dissertations & Theses Global. (305244249).	Unpublished dissertation
190	Stichter, J. P., Stormont, M., & Lewis, T. J. (2009). Instructional practices and behavior during reading: A descriptive summary and comparison of practices in title one and non-title elementary schools. Psychology in the Schools, 46(2), 172-183.	Focus is not on small groups
191	Strech, L. L. (1995). Ability grouping for elementary reading instruction and its relationship to the balanced literacy approach	Unpublished report
192	Taylor, B. M., Pearson, P. D., Clark, K., & Sharon Walpole. (2000). Effective schools and accomplished teachers: Lessons about primary-grade reading instruction in low-income schools. The Elementary School Journal, 101(2), 121-165.	Qual study- focus is not small groups
193	The Rector and the Board of Visitors of the University of Virginia. (2011). *PALS Electronic Lesson Plans.* Retrieved January, 2014, from http://pals.virginia.edu .	Not a study – describes formats for small groups
194	Tieso, C. (2005). The effects of grouping practices and curricular adjustments on achievement. Journal for the Education of the Gifted, 29(1), 60-89.	Focus on middle school and math groups
195	Tieso, C. L. (2003). Ability grouping is not just tracking anymore. Roeper Review, 26(1), 29-36.	Literature review
196	Tomlinson, C. A., & Association for Supervision and,Curriculum Development. (2001). How to differentiate instruction in mixed-ability classrooms. 2nd edition	Not a study
197	Torgesen, J. K. (2002). The prevention of reading difficulties. Journal of School Psychology, 40(1), 7-26.	Literature review
198	Tyner, B. (2009). Small-group reading instruction : A differentiated teaching model for beginning and struggling readers (2nd ed.). Newark, DE: International Reading Association.	Not a study – describes K-2 small group
199	Tyner, B., & Green, S. E. (2005). Small-group reading instruction: A differentiated teaching model for intermediate readers, grades 3-8 International Reading Association.	Not a study – describes 3-8 small groups
200	Vaughn, S., Hughes, M. T., Moody, S. W., & Elbaum, B. (2001). Instructional grouping for reading for students with LD: Implications for practice. Intervention in School and Clinic, 36(3), 131-37.	Focus is intervention
201	Vaughn, S., Linan-Thompson, S., Kouzekanani, K., Bryant, D. P., Dickson, S.,	Literature review –

	& Blozis, S. A. (2003). Reading instruction grouping for students with reading difficulties. Remedial and Special Education, 24(5), 301-315. doi:10.1177/07419325030240050501	focus on intervention
202	Vaughn, S., Moody, S. W., & Schumm, J. S. (1998). Broken promises: Reading instruction in the resource room. Exceptional Children, 64(2), 211-225.	Literature review
203	Vellutino, F. R., & Scanlon, D. M. (2002). The interactive strategies approach to reading intervention. Contemporary Educational Psychology, 27(4), 573-635.	Focus is on 1:1 intervention
204	Walpole, S., McKenna, M. C., Uribe-Zarain, X., & Lamitina, D. (2010). The relationships between coaching and instruction in the primary grades. Elementary School Journal, 111(1), 115-140.	Qual. Study- focus is not on small group
205	Weaver, C., & National Council of Teachers,of English. (1998). Practicing what we know: Informed reading instruction.	Online report
206	Wharton-McDonald, R., Pressley, M., & Hampston, J. M. (1998). Literacy instruction in nine first-grade classrooms: Teacher characteristics and student achievement. The Elementary School Journal, 99(2), 101-128.	Qual description – focus is not on groups
207	What, W. C. (2007). Success for all[R]. what works clearinghouse intervention report.What Works Clearinghouse.	Online review
208	What, W. C. (2009). Success for all[R]. WWC intervention report. (What Works Clearinghouse.	Online review
209	Wilkinson, I. A. G., & Townsend, M. A. R. (2000). From rata to rimu: Grouping for instruction in best practice new zealand classrooms. The Reading Teacher, 53(6), 460-471.	Qual. Description – no achievement measures
210	Wilkinson, L. C. (1988). Grouping children for learning: Implications for kindergarten education. Review of Research in Education, 15, 203-223.	Literature review
211	Wilks, R. (1996). Classroom management in primary schools: A review of the literature. Behaviour Change, 13(1), 20-32.	Literature Review
212	Wonder-McDowell, C., Ray Reutzel, D., & Smith, J. A. (2011). DOES INSTRUCTIONAL ALIGNMENT MATTER?: Effects on struggling second graders' reading achievement. Elementary School Journal, 112(2), 259-279.	Focus is on intervention /supplemental groups
213	Worthington, J. (1991). Reading groups: Problems, solutions	Not a study
214	Zakelj, A., & Grmek, M. I. (2013). Ability grouping and pupils' results on the national assessment of knowledge. Croatian Journal of Education-Hrvatski Casopis Za Odgoj i Obrazovanje, 15(2), 439-463.	Focus is on class composition