EXAMINING THE ROLE OF EXECUTIVE FUNCTION AS A MODERATOR OF A TIER 2 FIRST GRADE MATHEMATICS INTERVENTION

by

NICOLE A. HELLER

A DISSERTATION

Presented to the Department of Special Education and Clinical Sciences
and the Division of Graduate Studies of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2023

DISSERTATION APPROVAL PAGE

Student: Nicole A. Heller

Title: Examining the Role of Executive Function as a Moderator of a Tier 2 First Grade Mathematics Intervention

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Special Education and Clinical Sciences by:

Ben Clarke	Chairperson and Advisor
Nicole Giuliani	Core Member
Keith Smolkowski	Core Member
Nichole Kelly	Institutional Representative

and

Krista Chronister	Vice Provost for Graduate Studies

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded June 2023

DISSERTATION ABSTRACT

Nicole A. Heller

Doctor of Philosophy

Department of Special Education and Clinical Sciences

June 2023

Title: Examining the Role of Executive Function as a Moderator of a Tier 2 First Grade Mathematics Intervention

Low mathematics achievement in the United States has led to the recent advances in the development and evaluation of Tier 2 mathematics interventions designed to close gaps in mathematics that exist at school entry and prevent these gaps from widening. Because Tier 2 mathematics interventions are generally effective for improving mathematics outcomes for most, but not all, at-risk students, recent research has been focused on understanding for whom these interventions are most effective by examining student-level variables that may impact intervention effectiveness. One cognitive variable of particular interest has been executive function (EF).

Using extant data collected for the Fusion Efficacy Project (Clarke et al., 2015), this dissertation examined (1) whether there was differential response to the Fusion, a Tier 2 small-group first-grade mathematics intervention program, as a function of pre-intervention EF, measured by the Head Toes Knees Shoulders (HTKS) Task and (2) whether these differential effects varied by group size. Participants included first-grade students (n = 459) within classrooms (n = 53) randomly assigned to one of two treatment conditions (a small group of two or five students) or the business-as-usual control condition. Proximal and distal mathematics outcome measures were collected at pretest and posttest. Results of the moderation analyses indicated that, compared to the mathematics gains of students in the control condition, gains in mathematics of students in the intervention condition did not differ as a function of pre-intervention EF. However, moderation analyses did reveal a differential response based on group size for one of the four mathematics outcome measures, such that students with lower initial EF gained greater benefit from the intervention delivered in the smaller group (2:1) compared to the larger group (5:1). Results are discussed in the context of implications for future research and practice.

CURRICULUM VITAE

NAME OF AUTHOR: Nicole A. Heller

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon; Eugene, Oregon
University of California, Berkeley; Berkeley, California

DEGREES AWARDED:

Doctor of Philosophy, School Psychology, 2023, University of Oregon
Master of Science, Special Education, 2021, University of Oregon
Bachelor of Arts, Psychology, 2016, University of California, Berkeley

AREAS OF SPECIAL INTEREST:

Early mathematics intervention and assessment
The role of executive function in academic interventions
Multi-Tiered Systems of Supports (MTSS) in K-12 educational settings

PROFESSIONAL EXPERIENCE:

Graduate Research Fellow, School Psychology Program, University of Oregon, 2022, 2019-2020
Advanced Practicum Extern, HEDCO ADHD/LD Assessment Clinic, University of Oregon, 2021-2022
Graduate Teaching Fellow, School Psychology Program, University of Oregon, 2020-2022
School Psychology Advanced Practicum Student, Springfield Public Schools, Springfield OR, 2020-2021
School Psychology Practicum Student, Springfield Public Schools, Springfield OR, 2019-2020
Graduate Research Fellow, Center on Teaching and Learning, University of Oregon, 2019
Clinical Research Coordinator, Department of Developmental-Behavioral Pediatrics, Stanford Medical School, Stanford, California, 2016-2018
Research Assistant, Morrissey-Compton Education Center, Redwood City, California, 2016-2017

PUBLICATIONS:

Nelson, G., Park, S., Brafford, T., Heller, N. A., Crawford, A. R., & Drake, K. R. (2022). Reporting quality in math: Meta-analyses for students with or at risk of disabilities. *Exceptional Children, 88*(2), 125–144. https://doi.org/10.1177/00144029211050851

Wilkey, E. D., Shanley, L., Sabb, F., Ansari, D., Men, V., Cohen, J., Heller, N. A., & Clarke B. (2022) Sharpening, focusing, and developing: A study of change in nonsymbolic number comparison skills and math achievement in 1st grade. *Developmental Science*, *24*(3). https://doi.org/10.1111/desc.13194

Perrin, H. T., Heller, N. A., & Loe, I. M. (2019). School readiness in preschoolers with symptoms of attention-deficit/hyperactivity disorder. *Pediatrics*, *144*(2). https://doi.org/10.1542/peds.2019-0038

Loe, I. M., Heller, N. A., & Chatav, M. (2019). Behavior problems and executive function impairments in preterm compared to full term preschoolers. *Early Human Development, 130*, 87-95. https://doi.org/10.1016/j.earlhumdev.2019.01.014

ACKNOWLEDGMENTS

I would like to express gratitude to Ben Clarke for his guidance and support throughout the development and completion of this dissertation as well as his critical role as my academic advisor throughout my graduate school career. I would also like to express great thanks to Nicole Giuliani, Nichole Kelly, and Keith Smolkowski for serving as valuable members of my committee. I also wish to express great appreciation to Derek Kosty for his time and patience supporting my work. Additionally, I wish to express my deepest gratitude to my parents, Josh and Lori Heller, and a very special thank you to my amazingly supportive friends and cohort-mates, Taylor Lesner and Scott Rice. The research was supported by the Institute of Education Sciences, U.S. Department of Education, through grants R324A090341 and R324A160046 awarded to the Center on Teaching and Learning at the University of Oregon.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

I. INTRODUCTION

As part of the Fusion Efficacy Project Study (Clarke et al., 2015), Clarke and colleagues (in press) examined the overall intervention effects of Fusion, a Tier 2 first-grade mathematics intervention as well as whether these intervention effects varied between intervention groups with 2:1 and 5:1 student-instructor ratios. With regard to the main effects of the intervention, the authors found condition differences favoring the Fusion intervention on all four first-grade measures, ProFusion ($g = 0.77$), ASPENS ($g = 0.20$), TEMA-3 ($g = 0.07$), and easyCBM ($g = 0.02$), with the impacts for ProFusion and ASPENS statistically significant. When comparing between Fusion group size conditions, outcomes favored the 2:1 Fusion group on all four first-grade measures, ProFusion ($g = 0.16$), ASPENS ($g = 0.17$), TEMA-3 ($g = 0.21$), and easyCBM ($g = 0.29$), with the TEMA-3 and easyCBM reaching statistical significance.

Recently researchers have been interested in investigating moderators of differential response to mathematics interventions with the idea that it may lead to a more nuanced understanding of the student-level factors affecting intervention effectiveness (Fuchs & Fuchs, 2019; Miller et al., 2014). While other cognitive variables have been examined as potential moderators of mathematics intervention effects, this has yet to be done with Executive Function (EF). When designing the Fusion Efficacy Project Study (Clarke et al., 2015), the researchers also included the Head Toes Knees Shoulders (HTKS) task, a measure of EF, in the measurement net. The purpose of the present study was to examine EF as a moderator of response a Tier 2 first-grade mathematics intervention, as well as whether group size effects of the intervention are moderated by EF.

Consequences of Early Difficulties in Mathematics

One of the strongest predictors of later achievement across academic domains is mathematics knowledge at school entry (Duncan et al., 2007). Despite an increased emphasis for all students to demonstrate proficiency with mathematics, children who enter school with low mathematics skills continue to have difficulties for the rest of their school years (Jordan et al., 2006; Nelson & Powell, 2018). Mathematics achievement in early elementary school is significantly related to a range of outcomes including social-emotional development, mental health, and peer rejection (Morgan et al., 2011), as well as academic and vocational success, income, and socioeconomic status (SES) (Jimerson et al., 1999; Titz & Karbach, 2014). When examining the relationship between mathematics ability and SES, Ritchie and Bates (2013) found that mathematics ability at age seven was substantially and positively associated with SES at age 42, independently of relevant confounding variables, such as SES at birth and intelligence. In a systematic review of longitudinal studies of mathematics difficulties, Nelson and Powell (2018) found that early mathematics difficulties were predictive of mathematics achievement in later grades, and that, while students with mathematics difficulty made progress on mathematics measures, this progress still resulted in lower performance than that of students not experiencing mathematics difficulties.

According to the National Assessment of Education Progress (NAEP) results, many students are not meeting proficiency standards and students with disabilities are consistently behind their peers in fourth and eighth grade mathematics achievement. In 2019, 41 percent of fourth-grade students and 34 percent of eighth-grade students performed at or above the *NAEP Proficient* achievement level in mathematics. And for students with disabilities, even fewer performed at or above this proficiency standard (National Center for Educational Statistics, 2019). While the NAEP data captures student proficiency at fourth grade, we know that the development of mathematics learning difficulties (MLD) suggests that these difficulties typically start at school entry and continue over the course of students' academic careers. Such findings point to the need for early intervention in the area of mathematics to prevent early MLD from worsening over time and to mitigate the negative consequences with which they are associated (Frye et al., 2013).

Mathematics Interventions

Mathematics knowledge at school entry is an important predictor of later achievement across academic domains, and knowledge gaps in mathematics exist at school entry as school readiness varies greatly across students (Duncan et al., 2007). The existence of these knowledge gaps in mathematics lead to achievement gaps, which typically widen as students, both with and without disabilities, progress through elementary and middle school (Schulte & Stevens, 2015). Unfortunately, it has become increasingly clear that high quality Tier 1 instruction is not enough to close these mathematics achievement gaps in early elementary school (Clarke et al., 2016). Promisingly though, there have been recent advances in developing and evaluating Tier 2 mathematics interventions that focus on foundational number concepts and are designed to close these achievement gaps (Fuchs et al., 2021). These intervention programs typically target one domain of mathematics, such as whole number content, and employ an explicit and systematic instructional design framework with an emphasis on teacher modeling, opportunities to respond, and academic feedback (Gersten et al., 2009).

Dyson and colleagues (2011) evaluated a kindergarten number sense intervention designed to develop number competencies by targeting whole number concepts related to counting, comparing, and manipulating sets. Effects were examined using the Number Sense

Brief (NSB). The NSB assess student understanding of multiple critical early numeracy concepts including counting knowledge and principles, number recognition, number comparisons, nonverbal calculation, story problems and number combinations (Jordan et al., 2010). The researchers found statistically significant positive effects on the NSB at both immediate posttest (R^2 = .302) and delayed posttest six weeks later (R^2 = .299). Given the relative limited duration of the intervention, the impact on a measure targeting multiple early numeracy constructs indicated that targeted interventions in whole number were promising.

Clarke and colleagues (Clarke et al., 2011; Clarke et al., 2017; Doabler et al., 2019) examined the impact of ROOTS, a Tier 2 kindergarten mathematics intervention program, on the achievement of students at risk in mathematics. ROOTS is focused on developing whole number understanding in areas including (a) Counting and Cardinality: knowing number names and the count sequence, counting to tell the number of objects, and comparing numbers; (b) Number Operations: understanding addition as putting together and adding to, and subtraction as taking apart and taking from; and (c) Base 10/Place Value: working with numbers 11-19 to gain foundations for place value (CCSS-M, 2010). ROOTS is based on the concrete-representational-abstract model (CRA), an instructional sequence which involves first using concrete manipulatives, then 2D representational drawings, then standard numerals (Butler et al., 2003), and utilizes an explicit and systematic design framework, which is a highly structured method of teaching carefully sequenced material that includes teacher modeling with correct mathematical language and frequent practice opportunities with timely feedback (Gersten et al., 2009). As part of this framework, ROOTS uses scripted content delivery to promote high levels of implementation fidelity. Results indicated that students who received the intervention made significantly greater gains compared to students in the control group on the Test of Early

Mathematics Ability (TEMA-3), a standardized measure of early mathematics abilities focused on informal and formal number and operations knowledge.

Early intervention work has also extended to first grade. Fuchs and colleagues (2005) evaluated Number Rockets, a small group tutoring intervention designed to support first-grade students at risk for MLD. Based on the CRA model, the intervention focuses on teaching computation and building arithmetic fluency. The program emphasized features such as teacher modeling, opportunities to respond, and academic feedback within an explicit and systematic instructional design framework (Gersten et al., 2009). The authors found that the program, designed to teach conceptual understanding of mathematical ideas through the use of concrete models and manipulatives, created positive effects on both proximal and distal mathematics measures. In 2015, Gersten and colleagues (2015) conducted a replication of the Number Rockets program, implementing under more practical educational conditions. Results still indicated a positive effect on a distal measure of general mathematics achievement.

Bryant and colleagues (2008) evaluated the effects of a Tier 2 small-group first- and second-grade mathematics tutoring program aligned with the Texas Essential Knowledge and Skills standards. This program also utilized an explicit and systematic instructional framework alongside a concrete-semi-concrete-abstract (CSA) framework (Butler et al., 2003) to model and teach whole number concepts. The researchers found a significant intervention effect for second-grade students on the Texas Early Mathematics Inventories–Progress Monitoring (TEMI-PM) total standard score. In 2011, Bryant and colleagues examined the effects of an early numeracy preventative Tier 2 intervention on the mathematics performance of first-grade students with mathematics difficulties. The program focused on number and operation mathematical ideas and incorporated features of explicit and systematic instructional design such as error correction procedures, pacing, opportunities for meaningful practice, and a teaching routine consisting of modeling, guided practice, and independent practice. Results revealed that the students who received the intervention outperformed students in the comparison group on the three Texas Early Mathematics Inventories–Progress Monitoring (TEMI-PM) measures that focused on whole-number computation as well as the total standard score.

Clarke and colleagues (2014) also conducted a pilot study of Fusion, a Tier 2, small-group first-grade mathematics intervention program focused exclusively on whole number understanding. Results indicated that students who received the intervention made significantly greater gains compared to students in the control group on ProFusion, a proximal measure assessing conceptual understanding of whole-number content ($g = .82$).

Fusion Intervention

Fusion is a Tier 2 first-grade mathematics intervention intended to build students' proficiency with whole number critical concepts and skills. Fusion consists of 60 lessons, each 30 minutes in length, designed to be delivered over a period of 12 weeks. Each lesson covers mathematical content from two strands of the first-grade Common Core State Standards for Mathematics (CCSS-M, 2010) focused on whole number understanding: (a) Operations and Algebraic Thinking, which involves representing and solving problems involving addition and subtraction, understanding and applying properties of operations and the relationship between addition and subtraction, subtracting within 20, and working with addition and subtraction equations, and (b) Number and Operations in Base Ten, which involves extending the counting sequence, understanding place value, and using place value understanding and properties of operations to add and subtract. Fusion's scope and sequence introduces new concepts and skills in "tracks," with students practicing a variety of different skills each lesson. Activities within lessons build over time as increasingly advanced content is introduced. This sequencing allows for frequent review of previously taught content and supports students’ maintenance of mathematical skills.

The instructional architecture of Fusion is very intentional. The Fusion curriculum is carefully sequenced across content. First, each skill is taught to mastery and subsequently, the deliberate sequencing of content facilitates the priming of knowledge taught in previous lessons. The Fusion curriculum incorporates intentional scaffolding, in which interventionists provide students with less support over time and as students master the skills being taught. The first half of the curriculum, Lessons 1-30, focuses on number sense, basic number combinations, and place value concepts. In these first 30 lessons, students build proficiency with numbers up to 100 through identifying, modeling, writing, and sequencing numbers. Students are also explicitly taught strategies to fluently recall addition and subtraction number combinations within 10. The second half of the curriculum, Lessons 31-60, involves multidigit computation without regrouping and word problem solving. As lessons progress, the content becomes increasingly complex and gives students an opportunity to expand on skills taught earlier in the program. For example, Lessons 1-30 build place value understanding of two-digit numbers, then in Lessons 31-60, students learn to solve two-digit addition and subtraction problems and compare two-digit numbers using “greater than” and “less than” terminology. Additionally, in the second half of the program, students expand their repertoire of number combinations to include doubles facts and common number families (e.g., 3, 4, and 7). Lessons 31-60 are also designed to help students develop a solid understanding of mathematical problem-solving, which involves students learning the underlying structures of the word problem types identified in the first-grade CCSS-

M (2010), and how to represent and solve “add to,” “take from,” “put together,” and “take apart” problems.

Similar to ROOTS (Clarke et al., 2016) and Number Rockets (Fuchs et al., 2005), Fusion teaches mathematical models using a concrete-representational-abstract (CRA) framework intended to build students’ conceptual understanding (Butler et al., 2003; Gersten et al., 2009). For example, when learning place value of two-digit numbers, students use base-ten blocks and unit cubes (tens sticks and ones cubes) to model the tens and ones in a given number on a place value chart (concrete). Then, students draw pictorial representations of tens sticks and ones cubes to represent a given number (representational). Lastly, students represent the given number using standard numerals (abstract). In Fusion, the CRA framework is typically applied across multiple lessons, providing students with concurrent exposure to visual representations and mathematical symbols. Other mathematical models used in the program include number lines, number families, layered place value cards, and a hundreds chart.

To promote high-quality mathematics instruction and increase implementation fidelity, the Fusion intervention utilizes an explicit and systematic design framework (Gersten et al., 2009). As part of this, the Fusion intervention employs fully scripted lessons. These detailed scripts support teachers in (a) delivering clear demonstrations and explanations of targeted mathematics content, (b) facilitating frequent student practice opportunities, and (c) offering timely academic feedback. The lesson scripting facilitates teachers’ use of precise and consistent mathematical language and allows teachers to promote high-quality instructional interactions centered on whole number concepts and skills (Watkins & Solcum, 2014, p. 42). These interactions are intended to facilitate deep mathematical thinking and reasoning, through individual or group student mathematics verbalizations (Doabler et al., 2021; Fuchs et al., 2021).

For example, when teaching the commutative law of addition, the teacher writes two problems on the board (e.g., 3 + 1 =, 1 + 3 =) and asks students to discuss with their partner how the two problems are alike. Later on in the lesson, students practice explaining the commutative law to their partner using their own words.

As part of the explicit and systematic design framework, the Fusion intervention also involves active student participation during instruction, which includes group unison responses or choral responding. Choral responding allows all students to get high quality practice on all questions and allows the teacher to both assess the understanding of all students and maintain a positive learning environment by keeping all students on task. Because choral responding is only effective and successful when all students respond at exactly the same time (Watkins & Solcum, 2014, p. 45), the detailed scripts include explicit signals and cues to allow this to happen.

The intentional instructional design and delivery of the Fusion intervention greatly reduces the cognitive load placed on the learner and is hypothesized to meet the needs of all learners regardless of background knowledge or individual student characteristics such as initial skill or cognitive variables (Doabler & Fien, 2013; Shanley et al., 2021; Smith et al., 2016).

MTSS as an Approach to Address Low Mathematics Achievement

The intervention programs developed to date, including Fusion, were designed to be delivered in small groups of at-risk students within a Response to Intervention (RTI) or Multi-Tiered Systems of Support (MTSS) service delivery framework. RTI was originally conceptualized as an alternative way to determine eligibility for special education services (Individuals with Disabilities Education Act of 2004), that included a process of determining whether a student’s lack of growth was due to insufficient instruction or a learning disability. Since then, key characteristics of RTI have been adopted to support all students academically and behaviorally through an MTSS service delivery framework, instead of just students with learning disabilities (Fuchs & Vaughn, 2012; Vaughn & Swanson, 2015). MTSS is a framework for implementing instructional supports designed to meet students' demonstrated academic and behavioral needs and to improve academic and behavioral outcomes for all students (Fuchs & Vaughn, 2012; Schumacher et al., 2017). The MTSS framework involves instructional tiers that increase in instructional intensity. MTSS models typically consist of three tiers, the first tier consisting of core instruction in general education classrooms, the second tier consisting of small group supplemental instruction, and the third tier consisting of more intensive, often individualized, instruction (Gersten et al., 2009). In this three-tiered model in the context of mathematics, when a student does not adequately respond to Tier 1, or core, mathematics instruction, they will receive supplemental instruction at Tier 2 in the form of supplementary mathematics intervention that provides a more targeted learning experience. If the student does not respond to this supplemental instruction, they progress to Tier 3 where the intensity of the intervention is further increased.

One way to increase intervention intensity is decreasing the size of instructional groups as the severity of the academic deficit increases. For example, in a typical three-tiered MTSS model, the teacher to student ratio, or group size, in Tier 3 is smaller than group size at Tier 2 (Vaughn & Swanson, 2015). This smaller group size facilitates more opportunities for key teacher and student interactions hypothesized to support student acquisition of critical instructional content, including teacher models and individual student response opportunities (Baker et al., 2002; Gersten et al., 2009). However, because smaller intervention groups present substantial opportunity costs to schools (Clarke et al., 2020), it is important to explore the role of group size in intervention effectiveness.

In the reading literature, a meta-analysis on the impact of group size revealed larger effect sizes for smaller groups (Wanzek & Vaughn, 2007). Additionally, Vaughn and colleagues (2003) found that, when holding duration and instructional content constant, there were greater effects for smaller instructional groups (1:1 or 1:3 teacher to student ratio) compared to a larger instructional group (1:10 teacher-student ratio). However, the effects for the two smaller instructional groups (1:1, 1:3) did not differ significantly from each other.

In mathematics, Clarke and colleagues (Clarke et al., 2012-2016; Clarke et al., 2017; Doabler et al., 2019) examined the impact of group size on intervention effects for a kindergarten mathematics intervention, ROOTS, by randomly assigning treatment eligible students to one of three conditions: (a) ROOTS 2:1 group, (b) ROOTS 5:1 group, or (c) a no treatment control. While the authors found significant positive effects for ROOTS on both proximal and distal measures of mathematics achievement between treatment and control, they found no significant differences between the ROOTS 2:1 treatment group and the ROOTS 5:1 treatment group (Clarke et al., 2020). Clarke and colleagues (Clarke et al., in press) also examined the impact of group size on intervention effects for a first-grade mathematics intervention, Fusion, by randomly assigning treatment eligible students to one of three conditions: (a) Fusion 2:1 group, (b) Fusion 5:1 group, or (c) a no treatment control. The authors found significant positive effects for the treatment group on both proximal and distal measures of mathematics achievement between treatment and control, as well as significant differences between the Fusion 2:1 treatment group and the Fusion 5:1 treatment group favoring the 2:1 treatment group on two distal measures of first-grade mathematics achievement.

Understanding Response Variation: Differential Intervention Effects Based on Student-Level Variables

Modifications and efforts to improve intervention effectiveness do not automatically ensure that all students will respond. While Tier 2 mathematics interventions are generally effective for improving mathematics outcomes for most at-risk students (Bryant et al., 2008; Bryant et al., 2011; Clarke et al., 2020; Clarke et al., 2013; Dyson et al., 2011; Fuchs et al., 2005; Rolfhus et al., 2012), not all at-risk students respond to these interventions (Powell et al., 2017). More recently, researchers have been focused on understanding for whom these interventions are most effective and under what conditions by examining academic, cognitive, and behavioral student-level variables (Fuchs & Fuchs, 2019; Miller et al., 2014). Researchers note that investigations of differential response have the potential to lead to a greater understanding of the student-level factors affecting intervention effectiveness. Such insights would have implications for the designing of interventions and potentially enable greater specificity when deciding which students to target with particular interventions (Fuchs & Fuchs, 2019).

Recent studies have found differential intervention effects based on student-level variables such as initial mathematics skill or cognitive variables (Clarke et al., 2020; Fuchs et al., 2014, 2016, 2019; Powell et al., 2017; Shanley et al., 2021; Toll & Van Luit, 2013). When a student-level variable, such as a cognitive ability or academic skill, acts as a moderator of response, it means that the effect of intervention depends on the child's cognitive ability or academic skill at pretest. One cognitive variable of particular interest has been executive function (EF).

Executive Function

Executive function (EF) refers to higher order cognitive abilities used in planning, information processing, and problem solving for goal-directed behaviors in novel or challenging settings (Bierman et al., 2008). EF is thought to include three key constructs: working memory, cognitive flexibility (shifting), and inhibitory control (Best & Miller, 2010), which enable individuals to plan, organize, problem-solve, and manage their impulses (McClelland et al., 2014; Miyake et al., 2000). Working memory is the ability to hold, update, and manipulate information mentally, making it critical to planning and problem solving (Baddeley, 1986). Inhibitory control is the ability to suppress impulses and unwanted reactions and block out distractions to stay focused. Inhibitory control also allows individuals to think before acting in order to achieve a desired goal (Diamond, 2013). Cognitive flexibility involves the ability change or switch between tasks or mental sets and is important when adjusting to novel situations, taking on new perspectives, and discovering different ways to approach a problem (Diamond, 2013).

The Relationship Between Executive Function and Mathematics

Previous research suggests that there are relationships between the components hypothesized to make up EF and skills that contribute to early mathematics (Purpura, et al., 2016). Working memory allows students to hold multiple pieces of information in mind when solving a multi-step problem (e.g., number order, carrying values over in arithmetic operations; Fuchs et al., 2015; Purpura et al., 2016). Inhibitory control allows students to attend to the key pieces of information and ignore unnecessary information when solving problems, as well as applying correct strategies and inhibiting the use incorrect ones (Keller & Libertus, 2015; Purpura et al., 2016). Cognitive flexibility allows students to shift between different modalities of numeral representation, as well as different types of calculations, especially in multi-step problems (Lan et al., 2011; Purpura, et al., 2016).

Numerous studies have documented an association between EF and achievement in both reading and mathematics at a single point in time (Best et al., 2011; Clark et al., 2010; Morgan et al., 2019; Viterbori et al., 2015; Welsh et al., 2010). A number of longitudinal studies have demonstrated a clear predictive relationship between EF and academic achievement, where poor EF early in students' academic careers predicts later poor academic achievement. Measures of EF in early school-aged children are strong predictors of academic achievement throughout their academic careers (Arrington et al., 2014; McClelland et al., 2006; Nguyen & Duncan, 2018), even more so than standardized cognitive ability tests (Heckman & Kautz, 2012). For example, deficits in attention control in early childhood are associated with low academic achievement and a decreased likelihood of graduating high school (Rabiner et al., 2016). Furthermore, when examining the longitudinal relationship between children's cognitive characteristics and academic achievement, Sung and Wickrama (2018) found that children with higher levels of EF showed a faster growth rate in both reading and mathematics achievement.

More specific to mathematics, teacher and parent ratings of behavioral inattention are a significant predictor of later academic achievement, especially in mathematics (Duncan et al., 2007; McClelland et al., 2013). For example, teacher ratings of behavioral inattention can predict difficulties in conceptual and procedural aspects of mathematics (Raghubar et al., 2009). In a meta-analysis of contributions of students' initial mathematics, reading, and cognitive skills on subsequent mathematics performance, working memory was revealed to be a unique predictor of students' subsequent mathematics outcomes (Lin & Powell, 2021). It is suggested that EF plays a key role in the development of mathematics proficiency (Cragg & Gilmore, 2014). LeFevre and colleagues (2013) found that, among second- and third-grade children, executive attention, which includes executive function and executive working memory, predicted both mathematical knowledge and arithmetic fluency. However, executive attention only predicted growth in performance for mathematics fluency. In this study, the authors defined mathematical knowledge as knowledge of the number system (e.g., place value) and of arithmetic procedures (e.g., multi-digit addition) and arithmetic fluency as the speed of solutions to simple arithmetic problems.

Approaches To Measuring Executive Functioning

One widely used measure of EF is the Head Toes Knees Shoulder (HTKS) task. The HTKS is an observational assessment of behavioral self-regulation designed to measure multiple components of EF including working memory, inhibitory control, and cognitive flexibility (Ponitz et al., 2009; McClelland et al., 2014). The HTKS reports EF as a single score and unlike many other performance-based measures of EF, it integrates multiple EF components. McClelland and colleagues (2014) found that the HTKS was significantly related to the three key constructs involved in EF (cognitive flexibility, working memory, and inhibitory control) in both prekindergarten and kindergarten. Additionally, the HTKS is moderately to highly correlated with parent ratings of EF constructs, including attentional focusing and inhibitory control (Ponitz et al., 2009; McClelland et al., 2014), and teacher ratings of self-control and behavioral regulation (McClelland et al. 2007).

The HTKS measure has been used as a behavioral measure of EF in multiple studies examining the EF abilities of preschool and early elementary aged children. When examining the relationship between cognitive and behavioral measures of EF, Evers and colleagues (2016) found that the HTKS measure was highly correlated with cognitive EF measures (Digit Span backward, Block Recall, Day/Night Stroop, and Hearts & Flowers). Correlations between the cognitive EF measures and the HTKS were stronger when compared with correlations between the cognitive EF measures and the second behavioral measure, Tower (Kochanska et al., 1996), as well as correlations between the cognitive EF measures and teacher ratings of self-control and thoughtfulness. They hypothesized that these results might be because HTKS addresses all three EF components, but Tower primarily addresses inhibitory control. The HTKS has been shown to be moderately to strongly correlated with other established assessments of EF (Schmitt et al., 2017) and supports the concept of EF as a unitary construct (McClelland et al., 2021).

Using the HTKS to Examine the Relationship Between Executive Function and Mathematics

Several studies have used the HTKS as a behavioral measure of EF when examining the relationship between EF and early academic skills in preschool and early elementary school. Fuhs and colleagues (2014) used the HTKS measure as one of six measures of EF when exploring the longitudinal relationships between EF and early academic skills and found EF to be a strong predictor of mathematics gains in kindergarten. Schmitt and colleagues (2017) used the HTKS measure, along with three other behavioral measures of EF, when investigating the longitudinal relationships between EF and early academic skills. They found evidence that supported a longitudinal relationship between growth in EF and growth in mathematics.

Effects of EF Interventions on Mathematics Outcomes

Despite the relationship between EF and mathematics, most EF intervention studies focus on typically-developing children and do not feature academic achievement as an outcome of interest. Of the studies that do feature academic outcomes, few have demonstrated that the improvement in EF, attributed to the EF-focused intervention, transfers to academic outcomes. Instead, most meta-analyses have found that, despite these EF interventions leading to improved

EF, typically these effects do not transfer to academic achievement (Jacob & Parkinson, 2015; Melby-Lervåg et al., 2016; Melby-Lervåg & Hulme, 2013).

However, in a meta-analysis of attention training, Peng and Miller (2016) found the effects of attention training transferred to both cognitive and academic untrained tasks and that attention training was more effective for younger individuals. This supported findings from a previous meta-analysis that found that cognitive training, including working memory and attention training, leads to significantly more widespread transfer effects when applied to younger children (Wass et al., 2012). In general, studies that have demonstrated that growth in EF due to intervention transfers to academic achievement outcomes are limited to the preschool period. This suggests that the effects of EF interventions, especially interventions implemented after the preschool period, do not necessarily transfer to academics. This lack of transfer has led to a somewhat limited investigation of domain-specific EF interventions or combined EF and mathematics interventions on mathematics outcomes.

Effects of Combined EF and Mathematics Interventions on Mathematics Outcomes

Few studies have examined the effects of domain-specific EF interventions or combined EF and mathematics interventions on mathematics outcomes. When comparing domain-general working memory training, domain-specific working memory training, and a business-as-usual (BAU) control, Kroesbergen and colleagues (2014) found significant effects on mathematics outcomes for children who received either the domain-general or the domain-specific working memory training compared to the BAU control. Similarly, Nemmi and colleagues (2016) compared a working memory training, a number line training, and a combination working memory and number line training, finding that only the group that received the combined working memory and number line training improved significantly more than the control group.

Barnes and colleagues (2016) compared a preschool mathematics intervention plus attention training, the same preschool mathematics intervention without attention training, and a BAU control. The authors found that, while both the combined intervention and the mathematics intervention without attention training had positive effects on mathematics outcomes when compared to the BAU control, there were no significant differences between the two treatment groups on mathematics outcomes. When compared to the mathematics intervention without attention training, the combined intervention did have significant but small positive effects on attention. Due to the additional complexity and costs of adding EF components and implementing these combined interventions, it is important to gain a better understanding of how EF impacts mathematics interventions, which has yet to be done.

The Effect of Executive Function on Mathematics Learning and Intervention Outcomes

Recent longitudinal research suggests that EF at school entry may moderate growth in mathematics. Ribner and colleagues (2017) found that students with higher levels of EF are more likely to "catch up" to peers who perform better on assessments of early mathematics. Kindergarten students who enter school with both low EF and low mathematics skills show slower growth of mathematics skills over time compared to students who enter school with low mathematics skills but stronger EF. Ribner (2020) also found that kindergarten students differentially benefit from instruction depending on their EF abilities. Students with higher levels of EF benefit more from instruction than their peers with lower levels of EF. Additionally, Blair and McKinnon (2016) found that EF moderated the effect of preschool mathematics ability on mathematics ability at the end of kindergarten, where, among students with lower mathematics abilities in preschool, students with higher EF abilities achieved "higher than expected" levels of mathematics ability in kindergarten. However, outside of this small set of studies conducted in kindergarten, investigations of EF as a moderator of mathematics growth have been limited. Furthermore, there is no existing research examining EF as a moderator of mathematics intervention outcomes.

Differential Intervention Effects Based on Cognitive Variables

In recent years, researchers have examined the role of domain-general cognitive variables and response variation. Variables examined include working memory (Fuchs et al., 2014, 2016; Powell et al., 2017), processing speed (Powell et al., 2017), reasoning ability (Powell et al., 2017; Shanley et al., 2021), attention (Powell et al., 2017), language comprehension (Powell et al., 2017), visual-spatial skills (Shanley et al., 2021), and phonological skills (Shanley et al., 2021). Multiple components of EF, including working memory, attention, and reasoning (problem-solving), have been studied in the mathematics intervention research literature. This previous research focused on response variation by discrete constructs related to EF has suggested that EF as a whole construct may moderate the effects of mathematics interventions (Fuchs et al., 2014; Fuchs et al., 2016; Powell et al., 2017).

Shanley and colleagues (2021) explored the relations between mathematics achievement and domain-general cognitive skills, including fluid reasoning, visual-spatial skills, and phonological memory, in a sample of kindergarten students at-risk for MLD. Results indicated that all domain-general cognitive skills were modestly correlated with mathematics skill at pretest. However, the associations between these cognitive skills and mathematics achievement gains were not strong. Additionally, the authors did not find any differential effects in intervention response based on high or low domain-general cognitive skills.

Fuchs and colleagues (2014) investigated whether individual differences in working memory moderated effects of two variations of a fraction intervention for at-risk fourth-grade students, with one variation focusing on fluency and the other conceptual understanding. They found that working memory moderated the effects of the fluency versus conceptual condition on the number line measure. For students with very weak working memory, the conceptual variation of the intervention led to better performance, and for students with more adequate working memory, the fluency variation of the intervention led to better performance. And in 2016, Fuchs and colleagues examined whether individual differences in basic cognitive processes, including reasoning, working memory, and language comprehension, moderated the effects of two variations of a fraction intervention. In one variation, students were taught to provide "high quality" explanations when comparing fraction magnitudes or to solve fraction word problems. The other variation focused on word problems. Results indicated that, for students with weaker working memory, the explaining variation of the intervention was more effective, and for students with stronger reasoning ability, the word problems variation of the intervention was more effective.

Powell and colleagues (2017) investigated eight cognitive abilities and mathematics skills as potential moderators of responsiveness to intervention for two contrasting conditions, one focused on calculations and the other on word problems. The eight potential moderators included two pretest mathematics skills: (a) single-digit addition and subtraction fluency and (b) number line understanding, and six pretest cognitive abilities: (a) reasoning, (b) attentive behavior, (c) processing speed, (d) working memory-sentences, (e) working memory-counting, and (f) language comprehension. They found that there were two significant moderators of intervention effects for the calculation variation, including working memory-sentences and language comprehension. The effect of the calculation intervention was smaller for children with lower working memory as compared with children with more adequate working memory. The effect of the calculation intervention was smaller for children with lower language comprehension as compared with children with more adequate language comprehension. Lastly, they did not find any moderators of intervention effects for the word problems variation.

Study Goals and Research Questions

Given the emerging research base focused on cognitive variables and response variation to mathematics intervention and the well-documented relationship between EF and mathematics achievement, further investigation of EF as a moderator of mathematics intervention outcomes is warranted. Therefore, the overall purpose of this study is to better understand the role of EF in how students respond to mathematics intervention. The study utilized extant data from the Fusion Efficacy Project Study (Clarke et al., 2015). This study used a randomized control trial (RCT) design, blocking on classrooms where classrooms of students were randomly assigned to treatment (Fusion intervention) or control (standard district practice/BAU) conditions. This design allowed for an examination of the relationship between the HTKS and mathematics outcome measures, and EF (as measured by HTKS) as a moderator of the effects of the Fusion intervention.

Research Questions and Hypotheses

In addition to examining the descriptive statistics of each measure, including the HTKS, ProFusion, ASPENS, TEMA-3, and easyCBM Math, as well as examining correlations between all measures, I examined the HTKS as a moderator of Fusion intervention effects as well as group size differences in intervention effects. I hypothesized that EF would moderate the effects of Fusion, a Tier 2 mathematics intervention, such that the benefits of the Tier 2 mathematics intervention would be stronger for students with lower initial EF. Additionally, I hypothesized that there would be group-size differences in responsiveness based on EF, such that students with lower initial EF who received the intervention in a smaller group (2:1) would perform better on mathematics outcome measures compared to students with lower initial EF who received the intervention in a larger group (5:1).

II. METHOD

Research Design and Context

This study was conducted using extant data collected from the first two cohorts of a multi-year, four-cohort, federally funded efficacy project involving the Fusion intervention, a Tier 2 first-grade mathematics intervention. The Fusion Efficacy Project (Clarke et al., 2015) utilized a partially nested randomized controlled trial (Baldwin et al., 2011). Blocking on classrooms, 459 first-grade students were randomly assigned to one of three conditions: (a) Fusion small group with two students and one interventionist (2:1), (b) a Fusion large group with five students and one interventionist (5:1), and (c) a no-treatment control condition (i.e., BAU). In total, 91 students were assigned to the 2:1 Fusion condition, 230 students were assigned to the 5:1 Fusion condition, and 138 students were assigned to the no-treatment control condition. Students randomly assigned to the two treatment groups received the Fusion intervention in addition to district-approved core mathematics instruction. The first two cohorts resulted in a total of 92 Fusion intervention groups, including 46 2:1 Fusion intervention groups and 46 5:1 Fusion intervention groups. The methods for this study are further described in Doabler et al. (2021) and Clarke et al. (in press). The research team obtained Institutional Review Board approval for all study methods and procedures, and participants were treated in accordance with the ethical principles of the American Psychological Association.

Participants

Schools

Nine elementary schools from three Oregon school districts participated in the current study. One school district was located in the metropolitan area of Portland and two districts were located in suburban areas of western Oregon. Across the three districts, student enrollment ranged from 5,492 to 40,495 students. Within the 9 participating schools, 12% to 19% of students had disabilities, 11% to 27% were English learners, and 35% to 65% were eligible for free or reduced-price lunch. Less than 1% to 1% identified as American Indian or Native Alaskan, 1% to 16% as Asian, 1% to 3% as Black, 20% to 25% as Hispanic, less than 1% to 1% as Native Hawaiian or Pacific Islander, 48% to 67% as White, and 7% to 8% as more than one race.

Classrooms

The study took place in 53 total first-grade classrooms. All classrooms provided mathematics instruction in English and operated 5 days per week. Classrooms had an average of 21 students (*SD* = 5.6). The 53 classrooms were taught by 36 certified teachers. Of the 36 teachers, 89% identified as female and 86% as White, 3% Asian American or Pacific Islander, 3% American Indian or Alaskan Native, 6% two or more races, and 3% declined to respond. Teachers had an average of 13.2 years of teaching experience (*SD* = 9.1) and 7.4 years of first-grade teaching experience (*SD* = 6.4); 72% had a master’s degree in education; and 75% of teachers had completed an advanced mathematics course (i.e., calculus, algebra, statistics, or trigonometry) at the college level.

Students and Inclusion Criteria

In each participating classroom, all students with parental consent were screened in late fall of their first-grade year. The screening process consisted of three measures of the first-grade Assessing Student Proficiency in Early Number Sense battery (ASPENS; Clarke et al., 2011), a standardized measure of early mathematics proficiency. The three measures included (a) Magnitude Comparison, (b) Missing Number, and (c) Basic Arithmetic Facts and Base-10. Students who had an ASPENS composite score in the *Strategic* or *Intensive* categories based on winter benchmarks were considered eligible for the Fusion intervention and therefore considered at risk for MLD. Composite scores at or below the *Strategic* category suggest that students have less than a 50% chance of meeting end-of-year grade level expectations in mathematics (Clarke et al., 2011). A total of 1,075 first-grade students from Cohorts 1 and 2 were screened for Fusion eligibility.

Within each participating classroom, an independent evaluator rank ordered students with ASPENS composite scores in the *Strategic* or *Intensive* categories and then randomly assigned the 10 students with the lowest ASPENS composite scores to one of three conditions listed previously. In total, 1,075 students were screened, and 459 met the eligibility criteria. Classrooms with fewer than 10 eligible students were combined to form virtual randomization blocks. This procedure resulted in 47 randomization blocks from the 53 classrooms. Students were then randomly assigned within classrooms or virtual classrooms to the Fusion small group condition (*n* = 91), Fusion large group condition (*n* = 230), or the control condition (*n* = 138). Demographic data for the 459 Fusion-eligible students indicated that 16% received special education services, 17% were identified as English learners, and 53% identified as female. While the majority racial group of Fusion students identified as White (64%), 21% identified as Hispanic, 2% as Black, 3% as Asian, 1% as American Indian or Alaskan Native, 1% as Hawaiian or Pacific Islander, and 8% as more than once race. Student demographics are reported by condition in Table 1.

Table 1

Descriptive Statistics for Student Characteristics by Condition

	Fusion *n (%)*			Control *n (%)*
	2:1 Group Size	5:1 Group Size	Overall Fusion	
Total sample (*N*)	91	230	321	138
Sex				
Male	42 (46.2)	117 (50.9)	159 (49.5)	48 (34.8)
Female	48 (52.7)	110 (47.8)	158 (49.2)	87 (63.0)
Race				
American Indian or Alaskan Native	2 (2.2)	2 (0.9)	4 (1.2)	0 (0.0)
Asian	2 (2.2)	8 (3.5)	10 (3.1)	4 (2.9)
Black or African American	1 (1.1)	5 (2.2)	6 (1.9)	1 (0.7)
Hispanic or Latino	18 (19.8)	51 (22.2)	69 (21.5)	25 (18.1)
Hawaiian/Pacific Islander	1 (1.1)	3 (1.3)	4 (1.2)	0 (0.0)
White	62 (68.1)	136 (59.1)	198 (61.7)	94 (68.1)
Two or more	4 (4.4)	22 (9.6)	26 (8.1)	11 (8.0)
Special Education				
Eligible	14 (15.4)	40 (17.4)	54 (16.8)	21 (15.2)
Not Eligible	76 (83.5)	187 (81.3)	263 (81.9)	114 (82.6)
English Learner Status				
Yes	14 (15.4)	36 (15.7)	50 (15.5)	14 (10.1)
No	76 (83.5)	191 (83.0)	267 (83.2)	121 (87.7)
Home Language				
English	75 (82.4)	198 (86.1)	273 (85.0)	123 (89.1)
Other	15 (16.5)	29 (12.6)	44 (13.7)	15 (10.9)

Note. Due to missing data, not all percentages add up to 100.

Interventionists

Fusion intervention groups were taught by district-employed instructional assistants and by interventionists hired specifically for this study. A total of 39 interventionists taught the 92 Fusion groups. Among the interventionists, 94% identified as female. The majority of the interventionists were White (83%), with 3% identifying as Hispanic. The remaining 14% identified as another race or ethnicity or declined to respond. Most interventionists (72%) had previous experience providing small group instruction and 75% had a bachelor's degree or higher. Interventionists had an average of 5.3 years of teaching experience (*SD* = 7.2); 19% had a current teaching license; and 64% had taken an advanced mathematics course at the college level (e.g., calculus, algebra, statistics).

Procedures

Fusion Intervention

The Fusion intervention, as described above, was delivered to small groups of students (i.e., two or five students per interventionist), for 30 minutes a day, five days per week for approximately 12 weeks. Because Fusion is designed as a supplemental intervention, all students received Fusion in addition to core (Tier 1) mathematics instruction, and Fusion instruction occurred at times that did not conflict with core Tier 1 mathematics instruction. For all students, instruction started in early winter and ended in the spring. Implementation of the intervention began in early winter to provide students with the opportunity to respond to core mathematics instruction and therefore decrease the opportunity for the false identification of typically-achieving students during the screening process.

Professional Development

All interventionists participated in two 4-hour professional development workshops delivered by project staff. Interventionists attended the first workshop, which focused on content from Lessons 1-30, prior to the start of implementation of the Fusion intervention. Interventionists attended the second workshop, which focused on Lessons 31-60, after they had implemented approximately one quarter of the Fusion lessons. During the second workshop, interventionists had the opportunity to ask questions about the first half of the curriculum. Both workshops centered on validated practices in early mathematics instruction, small-group instruction, and classroom management. Staff leading the workshops explicitly modeled instructional practices, such as group response signals, immediate correction of student errors, and pacing of activities within lessons. Interventionists were provided opportunities to practice and receive feedback on lesson delivery from project staff as well as their peers. To promote implementation fidelity and enhance the quality of instruction, Fusion coaches conducted coaching visits which consisted of direct observations of lesson delivery, followed by feedback on instructional quality and fidelity of Fusion implementation. Interventionists received, on average, two coaching visits during Fusion implementation.

Fidelity of Implementation

Fidelity of Fusion implementation was measured via direct observation. Over the course of the intervention, trained research staff observed each Fusion group three times. Observers used a 4-point scale (1 = none, 2 = some, 3 = most, 4 = all) to rate the extent to which the interventionist (a) met the lesson's instructional objectives, (b) followed the lesson's teacher scripting, and (c) used the lesson's mathematics models. Observers also recorded whether the interventionist taught the number of activities given in the lesson.

Measures

Data collectors administered four student mathematics achievement measures at pretest (T_1) and posttest (T_2), all of which focused on critical whole number concepts and skills. Data collectors also administered one measure of EF at pretest (T_1). All measures were administered by trained research staff. Descriptions of each measure are provided below.

Student Mathematics Achievement

ProFusion (Clarke et al., 2014). ProFusion is a proximal measure aligned with the Fusion curriculum, developed by the Fusion research team to assess students' conceptual and procedural knowledge of basic whole number mathematics concepts. This measure is designed to be administered in an untimed, group setting. Students complete addition, subtraction, and story problems, as well as tasks such as writing numbers from dictation, writing numbers that match base ten block models, completing number sequences, and decomposing double digit numbers. Students also complete one-minute, timed addition and subtraction fluency measures. Criterion validity with other mathematics measures is $r = .56$ with the Early Numeracy Curriculum Based Measures (EN-CBM) and $r = .68$ with the Stanford Achievement Test, 10th Edition (SAT-10; Clarke et al., 2014).

Assessing Student Proficiency in Early Number Sense (ASPENS; Clarke et al., 2011). First-grade ASPENS assessment consists of four brief (1- to 2-minute), individually administered measures designed to assess student understanding of critical number concepts and ability to fluently use these concepts in mathematical performance tasks. These measures assess the ability to identify numerals (Numeral Identification), compare two numbers and determine which is greater (Magnitude Comparison), identify the missing number in a string of three numbers (Missing Number), and solve simple addition and subtraction computation problems that cross 10 (Basic Arithmetic Facts and Base 10). Test authors report test-retest reliability ranges from .70 to .90 across the four subtests. Criterion concurrent validity with the TerraNova 3 is reported as ranging from .51 to .63.

Test of Early Mathematics Ability – 3rd Ed. (TEMA-3; Ginsburg & Baroody, 2003). The TEMA-3 is a standardized, norm-referenced, individually administered measure of informal and formal number and operations knowledge. The TEMA-3 is designed to measure early mathematical ability, including skills related to counting, number facts and calculations, and related mathematical concepts among students ages 3 to 8 years 11 months. Test authors report alternate-form reliability of .97 and test-retest reliability ranges from .82 to .93. Concurrent validity with other criterion measures of mathematics is reported as ranging from .54 to .91.

easyCBM Math (Alonzo et al., 2006). EasyCBM Math is an online benchmark screening and progress monitoring system for kindergarten to eighth grade. Testing occurs on a secure web site and all test items are multiple choice. Reliability and validity of the assessments are well established. Internal reliabilities of first-grade easyCBM Math measures are high (.81-.84). Concurrent validity of easyCBM Math scores on the winter benchmark with the Stanford Achievement Test, Tenth Edition (SAT-10), ranges from .75 to .82. In the current study, the first-grade easyCBM measure was administered at posttest, and the second-grade version served as the follow-up assessment in second grade.

Executive Function

Heads, Toes, Knees, and Shoulders task (HTKS; McClelland et al., 2014; Ponitz et al., 2009). The HTKS is an observational assessment of behavioral self-regulation that measures a child’s working memory, inhibitory control, and cognitive flexibility. The HTKS is designed to be used with children ages 4 to 8 years and its expected administration time is 5 minutes. Interrater reliability is reported as .95 and the HTKS is positively correlated with parent ratings of attentional focusing (.25) and inhibitory control (.20), as well as teacher ratings of classroom behavioral regulation (.20; Ponitz et al., 2009).

Statistical Analysis

Clarke and colleagues (in press) examined overall effects of the Fusion intervention on each of the primary outcomes of mathematics achievement using a mixed-model (multilevel) Time × Condition analysis (Murray, 1998) designed to account for students partially nested within small groups (Baldwin et al., 2011; Bauer et al., 2008). Because an analysis of intervention effects is not what I am examining in the present study, I refer readers to Clarke et al. (in press) for a complete description of the analysis approach.

First, I examined the descriptive statistics of all measures of mathematics and EF as well as the assumptions of the statistical models. Then, to examine the relationship between the HTKS measure and measures of student mathematics achievement, Pearson's *r* bivariate correlations were estimated among the HTKS measure and all measures of student mathematics achievement, including ProFusion, ASPENS, TEMA-3, and easyCBM Math, at all available timepoints. Cohen (1992) defines small, medium, and large correlations as $r = .10$, .30, and .50, respectively, and descriptors of the strength of correlations are based on these definitions. The p-values for these correlations were not corrected for multiple comparisons as correlations were provided for descriptive purposes rather than inferential inference.

Next, to address my first primary research question regarding whether EF, as measured by the HTKS, moderates the overall outcomes of a Tier 2 mathematics intervention, I looked for differential responses to Fusion as a function of HTKS scores collected at pretest. I ran a moderation analysis to examine whether the Fusion intervention effects as measured by each measure of student mathematics achievement, including ProFusion, ASPENS, TEMA-3, and easyCBM Math, differed depending on students' initial EF, as measured by the HTKS. To estimate whether pretest EF moderated the effects of the Fusion intervention condition (small group variation) on measures of student mathematics achievement, I added an interaction term into the multiple regression model. I expanded the previous statistical models to include the predictor of differential response (HTKS) and its interaction with condition, time, and the Time × Condition term, resulting in a three-way interaction. The three-way Time × Condition × Pretest HTKS interaction provided an estimate of whether condition effects varied by initial EF.

Lastly, to address my second primary research question regarding whether EF, as measured by the HTKS, moderates differential outcomes of a Tier 2 mathematics intervention based on group size, I looked for differential responses to Fusion group size as a function of HTKS scores collected at pretest. I ran a moderation analysis to examine whether the group size intervention effects as measured by each measure of student mathematics achievement, including ProFusion, ASPENS, TEMA-3, and easyCBM Math, differed depending on students' initial EF, as measured by the HTKS. To estimate whether pretest EF moderated the effects of the Fusion 2:1 intervention condition as compared to the 5:1 intervention condition on measures of student mathematics achievement, I added an interaction term into the multiple regression model. I expanded the previous statistical models to include the predictor of differential response (HTKS) and its interaction with condition, time, and the Time × Group Size term, resulting in a three-way interaction. The three-way Time × Group Size × Pretest HTKS interaction provided an estimate of whether condition effects varied by initial EF.

Additionally, Type I errors associated with multiple comparisons for the two primary hypotheses (moderation analyses) were addressed using Benjamini-Hochberg's corrected p-values (p_{BH}; Benjamini & Hochberg, 1995) to control for the false discovery rate.

The data were analyzed using IBM SPSS Statistics (Version 26) predictive analytics software (IBM Corp., 2019). For the moderation analyses, I fit the statistical models to our data with full-information maximum likelihood estimation to minimize the potential for bias due to missing data (Allison, 2009; Graham, 2009). To deal with missing data, the Time × Condition and growth models were run with all available data on mathematics measures, but not with the HTKS measure of EF.

III. RESULTS

Descriptive Results and Baseline Equivalence

Descriptive Statistics of Measures and Analysis Assumptions

Table 2 presents the descriptive statistics, including means and standard deviations, medians, skewness, and kurtosis, by assessment time for each outcome measure except easyCBM, collected only at posttest as well as for the HTKS measure, collected only at pretest. On average from pretest to posttest, students gained 16.7 points (68%) on ProFusion, 19.6 points (89%) on ASPENS, and 6.9 points (20%) on TEMA-3.

The order of importance for regression assumptions are (1) independence, (2) heterogeneity of variance, and (3) normality (van Belle, 2008). The assumption of independence must be addressed through the partially nested analysis approach. Due to random assignment, in an RCT, there is no reason for the variances to differ across conditions. This is evidenced by the similar standard deviations across conditions presented in Table 3. Van Belle explains that "[f]or a two-sample test a three-fold difference in variances [SD 1.7 times as large] does not affect the probability of a Type I error" (p. 176). Therefore, hypothesis tests are unlikely to be affected by heterogeneity of variance unless the differences are very large (Box, 1953).

Finally, van Belle (2008) argues that normality is the least important of the assumptions. Through inspection of data plots and skewness and kurtosis statistics, nearly all measures at both assessment time points were normally distributed. Although strict criterion values do not exist for skewness and kurtosis, values greater than 1.0 were observed for the HTKS at pretest on skewness (-1.05) and kurtosis (1.17) and easyCBM total score at posttest on kurtosis (1.02). Skewness and kurtosis values from -1.0 to 1.0 or even larger likely do not meaningfully bias the regression-based result with samples of 100 or more (e.g., Wang & Thompson, 2007).

Table 2

Descriptive Statistics of Mathematics Measures and HTKS.

	T_1				T_2			
Measure	*Mean (SD)*	*Median*	*Skewness*	*Kurtosis*	*Mean (SD)*	*Median*	*Skewness*	*Kurtosis*
Profusion Composite	24.7 (9.5)	24.7	0.17	0.54	41.4 (12.8)	43.7	-0.47	-0.27
ASPENS Composite	22.1 (11.6)	23.0	-0.19	-0.95	41.7 (17.9)	42.0	0.21	0.83
TEMA-3	34.8 (7.2)	35.0	-0.34	0.41	41.7 (8.4)	41.0	0.04	0.72
easyCBM total score	-	-	-	-	25.3 (5.1)	26.0	-0.70	1.02
easyCBM percentile	-	-	-	-	34.4 (24.6)	30.0	0.59	-0.71
HTKS total score	40.5 (13.2)	42.0	-1.05	1.17	-	-	-	-

Table 3 presents the descriptive statistics, including means and standard deviations, by assessment time and intervention condition, Fusion 2:1, Fusion 5:1, overall Fusion Intervention, and control, for each outcome measure except the easyCBM Math, collected only at posttest as well as for the HTKS measure, collected only at pretest. On average from pretest to posttest, students in the 2:1 Fusion condition gained 21.0 points (83%) on ProFusion, 22.9 points (110%) on ASPENS, and 8.2 points (24%) on TEMA-3. Students in the 5:1 Fusion condition gained 18.9 points (77%) on ProFusion, 19.7 points (87%) on ASPENS, and 6.6 points (19%) on TEMA-3. Students in the either Fusion condition gained 19.5 points (79%) on ProFusion, 20.6 points (93%) on ASPENS, and 7.0 points (20%) on TEMA-3. Students in the control condition gained 10.3 points (42%) on ProFusion, 17.2 points (77%) on ASPENS, and 6.5 points (19%) on TEMA-3.

Table 3

Descriptive Statistics for Mathematics Measures by Condition and Assessment Time

		Intervention						Control	
		2:1 Group Size		5:1 Group Size		Overall Fusion			
Measure	Statistic	T_1	T_2	T_1	T_2	T_1	T_2	T_1	T_2
Profusion Composite	*M*	25.2	46.2	24.6	43.5	24.8	44.3	24.6	34.9
	(SD)	(9.5)	(10.7)	(9.4)	(12.3)	(9.4)	(11.9)	(9.7)	(12.3)
	n	90	85	224	214			137	130
ASPENS Composite	*M*	20.8	43.7	22.6	42.3	22.1	42.7	22.2	39.4
	(SD)	(11.4)	(16.6)	(11.7)	(17.4)	(11.6)	(17.2)	(11.55)	(19.37)
	n	91	85	230	214			138	130
TEMA Raw	*M*	34.7	42.9	35.0	41.6	34.9	41.9	34.6	41.1
	(SD)	(5.9)	(8.2)	(7.5)	(8.6)	(7.1)	(8.5)	(7.6)	(8.4)
	n	90	82	225	211			135	126
easyCBM total score	*M*	-	26.3	-	25.0		25.4	-	25.1
	(SD)		(4.6)		(5.4)		(5.2)		(4.8)
	n		70		182				111
easyCBM percentile	*M*	-	38.6	-	33.6		35.0	-	33.0
	(SD)		(24.2)		(24.7)		(24.7)		(24.5)
	n		70		182				111
HTKS total score	*M*	41.6	-	40.3	-	40.7		40.1	-
	(SD)	(12.0)		(13.4)		(13.0)		(13.7)	
	n	88		225				137	

Missingness and Differential Attrition

The potential for differential attrition for the Fusion intervention were presented in Clarke et al. (in press). Clarke et al. (in press) reported the overall rate of missingness at posttest to be 6.2% for the measures available at pretest, and the difference in rates of missingness among study conditions to be below 1.0% for posttest measures. The authors concluded that there were no statistically significant interactions between attrition and study condition predicting baseline outcomes, suggesting that the effect of attrition on outcomes would not likely threaten internal validity. The sample size for each measure at each timepoint is presented in Table 3.

Correlations

Table 4 presents correlations between measures with correlations below the diagonal representing pretest measures and correlations above the diagonal representing posttest measures. The correlations on the diagonal represent correlations between pretest and posttest for the same measure. All correlations were significant ($p < 0.01$) with the exception of the correlation between the posttest TEMA-3 and posttest easyCBM ($p < 0.05$). With regard to the relationships between the same measure at pretest and posttest, the relationships between pretest ProFusion and posttest ProFusion ($r = .63$), pretest ASPENS and posttest ASPENS ($r = .64$), and pretest TEMA-3 and posttest TEMA-3 ($r = .71$) were all large. Correlations between pretest mathematics measures including ProFusion, ASPENS, TEMA-3, and easyCBM were all large, ranging from $r = .68$ to $r = .75$. Similarly, correlations between posttest mathematics measures including ProFusion, ASPENS, and TEMA-3, were all large, ranging from $r = .53$ to $r = .68$, with the exception of the relationship between posttest ASPENS and easyCBM, which was medium ($r = .45$). Among mathematics measures, the correlations between easyCBM and other mathematics measures were the lowest, ranging from $r = .45$ to $r = .57$

Table 4

Correlations Between Measures with Pretest Below the Diagonal, Posttest Above the Diagonal and Pre-Post Correlations on the Diagonal

Measure	1	2	3	4	5
1. Profusion Composite	**.63****	.63**	.66**	.53**	.34**
2. ASPENS Composite	.68**	**.64****	.68**	.45**	.30**
3. TEMA-3	.75**	.73**	**.71****	.57*	.36**
4. easyCBM Post	.49**	.46**	.52**	–	.35**
5. HTKS Pre	.35**	.30**	.41**	.35**	–

Note. Correlations on the bottom diagonal are for pretest measures. Correlations on the top diagonal are for posttest measures. Correlations along the diagonal (in bold) are pre-post correlations. P-values for correlations were not corrected for multiple comparisons.

**$p<0.01$, *$p<0.05$

With respect to the measure of EF in this study, correlations between pretest mathematics measures (ProFusion, ASPENS, TEMA-3), and the HTKS, measured only at pretest, were medium, ranging from $r = .30$ to $r = .41$. Correlations between posttest mathematics measures (ProFusion, ASPENS, TEMA-3, easyCBM) and the HTKS, measured only at pretest, were medium, ranging from $r = .30$ to $r = .36$.

Table 5 presents correlations between the HTKS measure and all four posttest mathematics measures. Again, all correlations were significant ($p < 0.01$) with the exception of the correlation between posttest TEMA-3 and pretest HTKS ($p < 0.05$) and posttest ProFusion ($p > 0.05$), for the 2:1 Fusion condition. For the 2:1 Fusion condition, the correlations between pretest HTKS and each of the four posttest mathematics outcome measures were small to medium, ranging from $r = .21$ to $r = .35$. For the 5:1 Fusion condition, the correlations between pretest HTKS and each of the four posttest mathematics outcome measures were all medium, ranging from *r* = .31 to *r* = .45. For the overall Fusion condition (combined 2:1 and 5:1 group sizes), the correlations between pretest HTKS and each of the four posttest mathematics outcome measures were all medium, ranging from *r* = .31 to *r* = .39. Lastly, for the control condition, the correlations between pretest HTKS and each of the four posttest mathematics outcome measures were small to medium, ranging from *r* = .27 to *r* = .33.

When testing differences between the correlations between the HTKS measure and the four posttest mathematics measures, the only significant difference between any of the correlations was that between the 2:1 group and 5:1 group on ProFusion (*z* = 1.99, *p* = .023). This suggests that there may be an interaction between EF and group size on ProFusion (Hypothesis 1B formally tested below).

Table 5

Correlations Between Posttest Measures and HTKS by Condition

	Intervention			Control
Measure	2:1 Group Size	5:1 Group Size	Overall Fusion	
1. Profusion Composite Post	.21	.45**	.39**	.28**
2. ASPENS Composite Post	.30**	.31**	.31**	.27**
3. TEMA-3 Post	.26*	.40**	.37**	.33**
4. easyCBM Post	.35**	.37**	.37**	.33**

***p<0.01, *p<0.05*

Moderation of Overall Intervention Effects

My primary research question addressed whether EF, as measured by the HTKS, moderated the outcomes of Fusion, a Tier 2 mathematics intervention. Tables 6 and 7 present tests of differential response to the Fusion intervention on the four mathematics outcomes as a function of pretest HTKS scores. No statistically significant differences in mathematics outcomes at pretest were found, suggesting similar mathematics achievement at pretest by condition (Fusion vs. control).

The tests for moderation by the HTKS measure produced no statistically significant interaction effects on any of the four mathematics outcome measures. These tests of moderation suggest that intervention effects were not significantly moderated by pretest HTKS scores for any outcome: TEMA-3 ($t_{427} = 0.820$, $p_{BH} = .551$), ASPENS ($t_{439} = 0.255$, $p_{BH} = .798$), ProFusion ($t_{435} = 1.192$, $p_{BH} = .551$), or easyCBM ($t_{357} = 0.838$, $p_{BH} = .551$). See Table 6 and Table 7 for details. Thus, initial EF, as measured by the HTKS, did not affect the relationship between the Fusion intervention and mathematics outcomes.

Moderation Effects of Group Size

My secondary research question addressed whether EF, as measured by the HTKS, moderated differential outcomes of a Tier 2 mathematics intervention based on group size. Tables 8 and 9 present tests of differential response by Fusion intervention group size (2:1 vs. 5:1) on the four mathematics outcomes as a function of pretest HTKS scores. No statistically significant differences in achievement at pretest were found, suggesting similar mathematics achievement at pretest across group sizes (2:1 vs. 5:1).

Tests for moderation of group size effects by the HTKS produced no statistically significant interaction effects on any of the four mathematics outcome measures. These tests of moderation suggest that intervention differences by group size were not significantly moderated by pretest HTKS scores for outcomes including TEMA-3 ($t_{295} = -0.658$, $p_{BH} = .992$), ASPENS ($t_{302} = 0.115$, $p_{BH} = .992$), ProFusion ($t_{299} = -2.279$, $p_{BH} = .092$), and easyCBM ($t_{250} = -0.01$, $p_{BH} = .992$). However, these tests of moderation also suggest that, for the ProFusion measure, the moderation of intervention differences by group size by pretest HTKS scores was trending significant. This suggests that intervention differences by group size may be moderated by pretest HTKS scores for the ProFusion measure. See Table 8 and Table 9 for details.

Table 6

Results of Partially Nested Time × Condition Analyses That Tested Pretest HTKS as a Moderator of Differences in Gains in Mathematics Outcomes Between Fusion Students Nested Within Groups and Unclustered Control Students.

Effect or Statistic		TEMA-3	ASPENS	ProFusion
Fixed effects	Intercept	26.0**** (1.9)	12.0*** (3.8)	15.3*** (2.7)
	Condition	-0.2 (2.3)	-0.7 (4.6)	-0.6 (3.3)
	Time	7.2**** (1.7)	12.3*** (3.6)	9.7**** (2.4)
	Time × Condition	-1.0 (2.0)	2.2 (4.5)	5.7$^{\sim}$ (3.0)
	HTKS	0.2**** (0.1)	.3** (0.1)	.2**** (0.1)
	HTKS × Condition	0.0 (0.0)	0.0 (0.1)	0.0 (0.1)
	HTKS × Time	0.0 (0.0)	0.1 (0.1)	0.0 (0.1)
	HTKS × Time × Condition	0.0 (0.1)	0.0 (0.1)	0.1 (0.1)
Variances	Group-Level Intercept	7.4 (2.8)	21.4 (10.8)	13.3 (5.5)
	Group-Level Gains	1.0 (1.1)	2.9 (4.5)	2.2 (2.3)
	Student-Level Pre-Post Covariance	26.5**** (3.2)	47.3**** (6.2)	90.3**** (12.6)
	Residual	17.4**** (1.6)	38.1**** (3.3)	87.4**** (7.3)
	ICC	0.05	0.07	0.02
HTKS × Time × Condition effects	*p* value	.413	.798	.234
	BH *p* value	.551	.798	.551
	Standardized Coefficient and 95% CI	0.061 [-0.086, 0.208]	0.021 [-0.138, 0.179]	0.093 [-0.060, 0.246]
	Degrees of freedom	427	439	436

Note. Table entries show parameter estimates with standard errors in parentheses except for *p* values and degrees of freedom.

$^{\sim}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$. $^{****}p < .0001$.

Table 7

Results of Partially Nested Mixed-Model Analyses of Covariance on Posttest easyCBM Scores That Tested Pretest HTKS as a Moderator of Differences Between Fusion Students Nested Within Groups and Unclustered Control Students.

Effect or Statistic		easyCBM
Fixed effects	Intercept	13.8**** (1.7)
	Condition	-1.1 (1.5)
	HTKS	0.0**** (0.0)
	HTKS × Condition	0.0 (0.0)
Variances	Group-Level Intercept	0.1 (0.8)
	Residual	17.9**** (1.6)
	ICC	0.01
HTKS × Condition effects	*p* value	.402
	BH *p* value	.551
	Standardized Coefficient and 95% CI	0.077 [-0.103, 0.256]
	Degrees of freedom	357

Note. Table entries show parameter estimates with standard errors in parentheses except for *p* values and degrees of freedom.

$^{\sim}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$. $^{****}p < .0001$.

Table 8

Results of Fully Nested Time × Condition Analyses That Tested Pretest HTKS as a Moderator of Differences in Gains in Mathematics Outcomes Between Fusion Students in the 2:1 Condition and the 5:1 Condition Nested Within Groups.

Effect or Statistic		TEMA-3	ASPENS	ProFusion
Fixed effects	Intercept	25.5**** (1.5)	11.4**** (3.0)	14.2**** (2.1)
	Group Size (2:1)	1.8 (3.2)	-0.1 (6.1)	2.1 (4.3)
	Time	5.2**** (1.3)	14.0**** (2.9)	13.0**** (1.9)
	Time × Group Size	3.6 (2.7)	2.1 (6.0)	10.2* (4.0)
	HTKS	0.2**** (0.0)	.3**** (0.1)	.3**** (0.1)
	HTKS × Group Size	-0.1 (0.1)	0.0 (0.1)	0.0 (0.1)
	HTKS × Time	0.0 (0.0)	0.1* (0.1)	0.1*** (0.0)
	HTKS × Time × Group Size	0.0 (0.1)	0.0 (0.1)	-0.2* (0.1)
Variances	Group-Level Intercept	7.3** (2.8)	18.4* (9.3)	12.1* (5.2)
	Group-Level Gains	0.6 (0.9)	2.2 (4.3)	2.2 (2.1)
	Student-Level Pre-Post Covariance	25.9**** (3.3)	83.1**** (12.6)	45.5**** (6.3)
	Residual	16.8**** (1.6)	86.5**** (8.1)	35.9**** (3.4)
	ICC	0.03	0.02	0.06
HTKS × Time × Group Size effects	*p* value	.511	.908	.023*
	BH *p* value	.992	.992	.092$^{\sim}$
	Standardized Coefficient and 95% CI	-0.066 [-0.246, 0.132]	0.013 [-0.203, 0.228]	-0.255 [-0.474, -0.035]
	Degrees of freedom	295	302	299

Note. Table entries show parameter estimates with standard errors in parentheses except for *p* values and degrees of freedom.

$^{\sim}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$. $^{****}p < .0001$.

Table 9

Results of Fully Nested Mixed-Model Analyses of Covariance on Posttest easyCBM Scores That Tested Pretest HTKS as a Moderator of Differences Between Fusion Students in the 2:1 Condition and the 5:1 Condition Nested Within Groups.

Effect or Statistic		easyCBM
Fixed effects	Intercept	19.3^{****} (1.1)
	Group Size (2:1)	1.1 (2.4)
	HTKS	0.1^{****} (0.0)
	HTKS × Group Size	0.0 (0.0)
Variances	Group-Level Intercept	0.0 (0.0)
	Covariance	10.9^{****} (0.0)
	Residual	12.5^{****} (2.1)
	ICC	0.0
HTKS × Group Size effects	*p* value	.992
	BH *p* value	.992
	Standardized Coefficient and 95% CI	-0.002 [-0.322, 0.319]
	Degrees of freedom	250

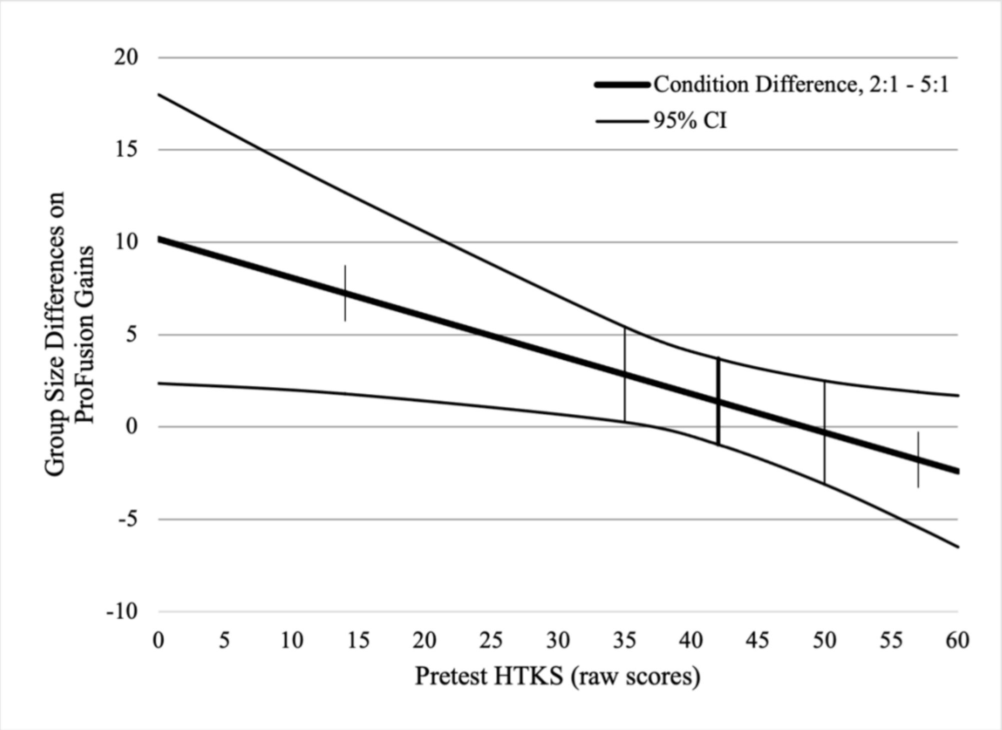
Note. Table entries show parameter estimates with standard errors in parentheses except for *p* values and degrees of freedom.

$^{\sim}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$. $^{****}p < .0001$.

Although the moderation effect was not significant for ProFusion ($p_{BH} = .092$), prior to correcting for multiple comparisons the moderation effect was significant ($p = .023$). To illustrate this relationship, figure 1 presents condition differences in ProFusion outcomes across the range of pretest HTKS scores. The figure illustrates the estimated difference between conditions with the 95% confidence intervals across the range of pretest HTKS scores (Preacher et al., 2006). Zero on the vertical axis represents no difference between conditions. The vertical lines within the graph represent sample percentiles (5th, 25th, 50th, 75th, and 95th). The vertical lines show that about 50% of the students had pretest HTKS scores below 42, 25% below 35, and 5% below 14. The confidence bounds exclude zero at values of 0 to 37, or for students scoring between the 0th and 36th percentile for the sample on the pretest HTKS assessment. This suggests that students scoring in this range on the HTKS achieved greater gains on ProFusion when receiving the Fusion intervention in the 2:1 intervention condition compared to students receiving the Fusion intervention in 5:1 intervention condition.

Figure 1

Differential Effects of Group Size on ProFusion Gains Based on Pretest HTKS Scores



The vertical axis shows the difference between group sizes—zero on the vertical axis represents no difference between group sizes—and the horizontal axis represents the range of pretest HTKS scores. The heavy decreasing line depicts the mean difference between group sizes at each pretest value. The two thinner, outer lines show the 95% confidence interval around the mean estimate. To show the location of the sample on the graph, the vertical lines within each figure depict the median (heavier vertical line), 25th and 75th percentiles (thinner long lines), and the 5th and 95th percentiles (short outer lines). For example, a score of 15 represents the lower 5th sample percentile at pretest.

IV. DISCUSSION

Developing mathematics proficiency in early elementary school is critical to later academic and life success, as early mathematics difficulties are predictive of poor academic achievement in later grades (Nelson & Powell, 2018). Therefore, the development of effective early mathematics interventions designed to close achievement gaps is critical (Frye et al., 2013; Gersten et al., 2009). As the field of early mathematics interventions progresses, researchers have begun to use student-level variables to better understand for whom academic interventions are most effective and under what conditions as potential avenue to increase intervention effectiveness and reduce nonresponse (Fuchs & Fuchs, 2019; Miller et al., 2014). The results of the current study contribute to the literature by examining EF as one student-level variable that may impact mathematics intervention outcomes. EF is an important variable to explore because the current research base has documented relationships between the components hypothesized to make up EF – working memory, inhibitory control, and cognitive flexibility – and the skills that contribute to early mathematics understanding (Purpura, et al., 2016). Furthermore, the literature documents a clear association between EF and mathematics achievement (Best et al., 2011; Morgan et al., 2019) as well as clear predictive relationships between the two constructs. Given this well-documented relationship between EF and early mathematics, EF is a plausible moderator of response to mathematics intervention. While previous studies have examined other cognitive abilities, including components of EF, as moderators of intervention response, no prior studies have investigated the relationship between overall EF and intervention response.

Therefore, the purpose of this study was to investigate the relationship between students' initial EF and their response to a Tier 2 first-grade mathematics intervention. In the following sections, I summarize and interpret key study findings, note limitations to the current study, and suggest future directions for research.

Summary and Interpretation of Results

RQ1A: Moderation of Overall Intervention Effects by EF

Recent studies have investigated the role of domain-general cognitive variables on response variation for mathematics interventions. Multiple components of EF, including working memory (Fuchs et al., 2014, 2016; Powell et al., 2017), attention (Powell et al., 2017), and reasoning or problem-solving (Powell et al., 2017; Shanley et al., 2021), have been studied in the mathematics intervention research literature. Results from these studies are inconsistent with regard to which domain-general cognitive variables moderate mathematics intervention response. However, no prior studies have investigated EF as a unitary construct related to differential mathematics intervention response. Therefore, the primary goal of this study was to investigate whether EF, as measured by the HTKS, moderates the outcomes of the Fusion intervention. Due to the well-documented relationship between EF and mathematics and the instructional architecture of the Fusion curriculum, it was hypothesized that EF would moderate the impact of Fusion, such that students with lower initial EF would show stronger gains in response to the Fusion intervention. Contrary to this hypothesis, analyses yielded no statistical evidence for moderation by EF, suggesting that the impact of the Fusion intervention was comparable for at-risk students with a wide range of pre-intervention EF profiles.

RQ1B: Moderation of Group Size Effects by EF

A secondary goal of this study was to investigate whether EF, as measured by the HTKS, moderates differential response to the Fusion intervention based on group size. It was hypothesized that there would be a differential response to differences in group size based on EF such that students with lower initial EF who received the Fusion intervention in a smaller group (2:1) would benefit more from the Fusion intervention compared to students with lower initial EF who received the intervention in a smaller group (5:1). Contrary to my hypothesis, analyses yielded no statistically significant evidence for moderation of group size effects by EF for any of the four mathematics outcome measures. However, on the ProFusion measure, analyses revealed a moderation effect by EF that trended toward statistical significance (p = .023; p_{BH} = .092), such that students with lower initial EF benefitted more when Fusion was delivered in a smaller group (2:1) format. Importantly, this potential moderation effect was only present for one outcome measure, ProFusion (a proximal but broad measure of skills taught in the Fusion intervention), whereas analyses indicated that intervention differences by group size were not moderated by students' initial EF across remaining outcome measures including TEMA-3, ASPENS, and easyCBM. Further analyses suggest that students scoring between the 0th and 36th percentile (corresponding to raw scores from 0 to 37) on the HTKS achieved greater gains on the ProFusion when receiving the Fusion intervention in the 2:1 intervention condition compared to the 5:1 intervention condition. This finding of potential differential response based on group size for the ProFusion mathematics suggests that, when implementing the Fusion intervention, smaller group sizes may lead to greater gains in mathematics for students with weaker initial EF. However, any of interpretation of this finding should be done with caution, as the effect of differential response based on group size was only statistically significant prior to correcting for multiple comparisons.

Implications and Future Directions

Findings from this study point to several interesting implications to consider as we attempt to better understand student-level variability in response to effective mathematics intervention. Despite previous findings suggesting that EF might impact intervention outcomes (Blair and McKinnon, 2016; Fuchs et al., 2014; Powell et al., 2017; Ribner, 2020), results suggested that the impact of the Fusion intervention was similar for at risk students regardless of pre-intervention EF profiles. EF is a higher order construct that is comprised of and can be broken down into three distinct cognitive constructs – working memory, inhibitory control, and cognitive flexibility. Therefore, it is possible that in the present study, no moderation effects were detected because intervention effects were only moderated by one of these three constructs. Because the HTKS reports student EF as a single score, this study was unable to examine whether one of the three components of EF may have acted as a moderator of overall intervention effects. While some research has suggested that EF can best be explained as a unitary construct in the context of the relationship between EF and mathematics in the early elementary grades (Brydges, et al., 2012; Nugyen, et al., 2019; Wiebe et al., 2011; Willoughby, et al., 2012), other researchers have suggested that a multi-factor model of EF may be more appropriate (Lerner & Lonigan, 2014). Therefore, future studies should utilize a measure of EF that reports individual scores for each of the three components of EF or utilize multiple measures to individually assess all three EF components.

Additionally, future studies should investigate whether subcomponents of EF moderate mathematics intervention effects. For example, in a meta-analysis, Lin and Powell (2021) found that working memory was a unique predictor of students' subsequent mathematics outcomes. Additionally, multiple studies have found that working memory can moderate the effects of different intervention conditions, such that one variation of the intervention is significantly more effective for students with weaker working memory (Fuchs et al., 2014; Fuchs et al., 2016). Future studies should continue to explore working memory as a potential moderator of mathematics intervention. More broadly, future research should continue to explore the overall relationships between each EF component (working memory, inhibitory control, cognitive flexibility) and mathematics learning as this may offer insight into how different EF components impact mathematics development and mathematics intervention response.

A better understanding of student-level response variation may also help researchers find ways to increase the overall efficacy of their interventions. Therefore, another possible explanation for the lack of differential response observed in the current study lies in the instructional architecture of the Fusion intervention. Fusion incorporates explicit and systematic instructional design features to promote accessibility to at-risk students with a wide range of learning needs and individual characteristics including pre-intervention EF. The lack of moderation in the present study may therefore lend support to the theory that mathematics interventions can be designed to support students with MLD in compensating for underlying cognitive deficits (Fuchs et al., 2014). Because of the clearly established relationship between EF and mathematics, researchers have suggested addressing EF to improve mathematics by way of EF training in isolation (Jacob & Parkinson, 2015; Melby-Lervåg et al., 2016; Melby-Lervåg & Hulme, 2013) as well as domain-specific EF training (Barnes et al, 2016; Krosenbergen et al., 2014; Nemmi et al., 2016). However, a simpler solution may be to apply the deliberate instructional architecture and design and delivery of the Fusion intervention to the design of future mathematics interventions (or modifications of existing interventions), to meet the needs of students with a wide range of EF skills. Future research should examine whether effects of interventions that vary in the extent to which they include or exclude the instructional design features found in the Fusion intervention are moderated by EF. For example, I hypothesize that if this study were to be replicated in the context of an early mathematics intervention that differed from Fusion in instructional architecture, results may support differential intervention effects based on student EF profiles. Similarly, it would be interesting to contrast moderation by EF effects with students randomized to receive either an intervention with a robust instructional architecture, such as the Fusion intervention, or an alternate early mathematics intervention that lacks the design features hypothesized to negate the impact of EF on instructional response.

While the simplest solution for researchers and educators may be to design or select interventions with these key instructional design features, there is still value in exploring other ways to utilize knowledge of the relationship between EF and mathematics within mathematics intervention research. Previous research has revealed mixed results regarding (a) whether improvement in EF from EF interventions (e.g., cognitive training) transfers to mathematics outcomes (Jacob & Parkinson, 2015; Melby-Lervåg et al., 2016; Melby-Lervåg & Hulme, 2013; Peng & Miller, 2016; Wass et al., 2012) and (b) whether combined EF and mathematics interventions (e.g., domain-specific cognitive training) support greater mathematics gains (Barnes et al, 2016; Krosenbergen et al., 2014; Nemmi et al., 2016). Because this area of research has produced mixed results, further inquiry is warranted. Additionally, previous research has examined the addition of specific intervention components to help compensate for students' cognitive skill limitations (Fuchs et al., 2016). Furthering this line of research by isolating and studying which specific components of systematic and explicit design are most effective at helping students compensate for deficits in EF is also an important direction for continued research.

This study's results examining group size and EF suggest that students with lower initial EF who were assigned to smaller (2:1) intervention groups made greater gains compared to students with lower initial EF who received the intervention in larger groups (5:1) on one of four outcomes measures. This effect of differential response based on group size was only found on the ProFusion outcome measure and was only statistically significant prior to correcting for multiple comparisons (p = .023; p_{BH} = .092). ProFusion is a research-developed measure closely aligned with the Fusion curriculum. ProFusion was designed to assess students' conceptual and procedural knowledge of basic whole number mathematics concepts and includes tasks such as writing numbers from dictation, writing numbers that match base ten block models, completing number sequences, and decomposing double digit numbers as well as addition, subtraction, and story problems. It also includes a one-minute, measure of addition and subtraction fact fluency. Of the four mathematics outcome measures, ProFusion was most proximal to the intervention, such that items on this measure were very similar to activities students completed as part of the intervention. The characteristics of the ProFusion measure may offer an explanation as to why there was a potential group size moderation effect by EF. Compared to the other outcome measures, ProFusion does not require as much generalization or application of the mathematics skills taught as part of the Fusion intervention. Over the course of the intervention, students received substantial practice with the types of tasks included in the ProFusion measure. Because of this, ProFusion also measured student mastery of the problem-solving procedures and approaches taught in the Fusion intervention. That being said, it should be noted that ProFusion is still a broad measure of first-grade mathematics whole number standards (CCSS-M, 2010). As such, a potential moderation on an outcome measure that reflects the comprehensive nature of the whole number knowledge students are expected to understand and apply in first grade is significant from both a clinical and practical standpoint.

A possible explanation for the potential group size moderation results is that a reduction in the student to teacher ratio facilitates greater individualization of instruction and more opportunities for active learning, increasing in the overall intensity of the intervention (Foorman & Torgensen, 2001; Mellard et al., 2010). It could be hypothesized that students with EF difficulties will receive greater benefit from the individualization of instruction and increased opportunities for active learning compared to peers with average EF. Importantly however, because the effect of differential response based on group size was only statistically significant prior to correcting for multiple comparisons, interpretation related to this finding should be done with caution. While research has investigated the relationship between instructional interactions and student outcomes (Doabler et al., 2019) including with this data set (Doabler et al., 2021), no prior studies have examined the relationship between EF, group size, and intervention intensity. Future studies should investigate whether and how intervention group size and instructional intensity variables affect mathematics gains among students with low EF.

Researchers have suggested including cognitive measures in the screening process to help assign students to interventions to which they are most likely to respond adequately (Powell et al., 2017). However, the findings of the current study suggest that, for practitioners, it would generally be more efficient to select an intervention that works equally well for students regardless of their initial EF skills. It should be noted that research findings related to group size complicate this general suggestion. Since a trend was detected (i.e., greater benefit for students with lower EF in the smaller group size) researchers should continue to investigate questions of this type but be cautious in advocating for modifications to current screening practice in schools.

This study was complicated by the lack of consensus on the true definition of EF and how to measure EF. In particular, there is some disagreement about whether and how EF differs from effortful control (EC). Liew (2011) claims that EF and EC are complimentary constructs that both fall under the umbrella of self-regulation. According to Zhou and colleagues (2011), EF and

EC are both related to self-regulation, but have each been the focus of research in different fields. EC is the subject of temperament research and EF is more often the focus of cognitive neuroscience and psychology research. Nigg (2016) hypothesizes that EC emerges from and maps onto EF, and that both EF and EC are “top-down” aspects of self-regulation, but EC is narrower than EF. McClelland and colleagues (2021) explain that behavioral self-regulation is the use of EF skills in different situations. As this study was done in the context of education and school psychology, I focused on EF as my construct of interest. However, while future studies should continue to explore the broader relationship between EF and mathematics learning, they should also consider examining the relationships between mathematics learning and constructs similar to EF, such as EC.

Limitations and Future Directions

Interpreting these findings should be made in the context of the study’s limitations. The HTKS reports EF as a single score, with no subscores reflecting the three individual components of EF. Because of this, this study was unable to investigate whether any of the three components of EF individually moderated intervention effects. As suggested previously, future studies should use a measure of EF that reports separate scores for each of the three components of EF or multiple measures. Additionally, among this sample, the variability seen in HTKS scores was slightly less than (*SD* = 13.2), but still comparable to that observed in previous studies examining the psychometric properties of the HTKS, with standard deviations ranging from 15 to 19 (McClelland et al., 2014). Therefore, it is possible, but not likely, that the presence of reduced variance could have limited the ability to detect a significant moderation effect. It is important to note that the HTKS is a behavioral measure of EF and studies have suggested that behavioral or performance-based measures and parent or teacher rating measures of EF assess different underlying cognitive constructs (Toplak et al., 2012). Therefore, when measuring EF in future studies, researchers should prioritize the use of multiple measures of EF, ideally including a combination of both behavioral measures of EF and parent or teacher ratings of EF. Additionally, researchers should continue to look at the relationship between different measures of EF, including behavioral measures such as the HTKS and parent- or teacher-reported measures and early mathematics performance.

Another limitation of this study is that EF was only measured at pretest. Therefore, this study was unable to examine whether the Fusion intervention had any impact on students' EF. Previous research has suggested that there is a bidirectional relationship between EF and mathematics (Fuhs et al., 2014; Hassinger-Das er al., 2014; McKinnon & Blair, 2019). However, none of the existing research examines this relationship in the context of intervention. Future studies exploring the relationship between EF and mathematics in an intervention context should measure EF at both pretest and posttest as a way to examine the possibility of a bidirectional relationship. Of interest would be examining whether mathematics interventions support the growth of EF. In addition, studies can investigate whether particular types of mathematics interventions or interventions focused on specific areas of mathematics also promote growth in EF. While many studies have focused on the development of EF during preschool, substantial development is seen in all EF components through adolescence (Best & Miller, 2010). This finding and research documenting different developmental trajectories for EF components, suggest future studies exploring longitudinal growth patterns of both EF and EF components and their interactions with mathematics interventions at different time points are warranted.

Another limitation of this study was that the four mathematics outcome measures selected for this study were all highly correlated with each other. A latent variable of general mathematics achievement could have been created for use in the moderation analyses. However, while the correlations between the mathematics measures were moderate to strong (r = .46 to r = .75), the individual measures were designed to assess specific constructs within mathematics. In addition, prior studies using the Fusion Efficacy Project data examined intervention effects on each individual mathematics outcome measure (Clarke et al., in press). For these reasons, I choose to examine each mathematics outcome measure individually.

Finally, it is important to note that the findings and implications from this study cannot be generalized to other early mathematics interventions. The implications discussed previously are specific to the Fusion intervention. As suggested above, it is critical that future research continues to investigate research questions related to differential response with a range of intervention types, measures of EF, and at multiple time points.

Conclusion

Recently intervention research has started to explore student-level moderators of responsiveness. These variables have the potential to facilitate more effective educational decision-making by providing insight regarding how student-level characteristics, including cognitive variables such as EF, may impact the appropriateness of an intervention for a particular student (Miller et al., 2014). This study extends the literature in this area by examining EF as one such student-level moderator of intervention response. Findings revealed that there was no differential response to the Fusion intervention based on initial EF. However, one mathematics outcome measure indicated a potential differential group size effects based on initial EF, such that students with lower initial EF receiving the intervention in smaller groups (2:1) made greater gains than those receiving the intervention in larger groups (5:1).

The findings of this study have implications for a range of audiences including researchers, curriculum developers, and educators including those responsible for selecting and implementing intervention curricula. This study provides considerations for the above groups of people when deciding how best to design and deliver mathematics interventions in the context of an MTSS framework to support the mathematics learning of all students.

REFERENCES CITED

Allison, P. D. (1990). Change scores as dependent variables in regression analysis. *Sociological Methodology*, *20*, 93. https://doi.org/10.2307/271083

Alonzo, J., Tindal, G., Ulmer, K., & Glasgow, A. (2006). *easyCBM online progress monitoring assessment system.* University of Oregon. http://easycbm.com

Arrington, C. N., Kulesz, P. A., Francis, D. J., Fletcher, J. M., & Barnes, M. A. (2014). The contribution of attentional control and working memory to reading comprehension and decoding. *Scientific Studies of Reading*, *18*(5), 325–346. https://doi.org/10.1080/10888438.2014.902461

Baddeley, A. (1986). *Working memory*. Clarendon Press/Oxford University Press.

Baker, S., Gersten, R., & Lee, D.-S. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *The Elementary School Journal*, *103*(1), 51–73. https://doi.org/10.1086/499715

Baldwin, S. A., Bauer, D. J., Stice, E., & Rohde, P. (2011). Evaluating models for partially clustered designs. *Psychological Methods*, *16*(2), 149–165. https://doi.org/10.1037/a0023464

Balu, R., Zhu, P., Doolittle, F., Schiller, E., Jenkins, J., & Gersten, R. M. (2015). *Evaluation of response to intervention practices for elementary school reading* (NCEE No. 2016-4000). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://files.eric.ed.gov/fulltext/ED560820.pdf

Barnes, M. A., Klein, A., Swank, P., Starkey, P., McCandliss, B., Flynn, K., Zucker, T., Huang, C.-W., Fall, A.-M., & Roberts, G. (2016). Effects of tutorial interventions in mathematics and attention for low-performing preschool children. *Journal of Research on Educational Effectiveness*, *9*(4), 577–606. https://doi.org/10.1080/19345747.2016.1191575

Bauer, D. J., Sterba, S. K., & Hallfors, D. D. (2008). Evaluating group-based interventions when control participants are ungrouped. *Multivariate Behavioral Research*, *43*(2), 210–236. https://doi.org/10.1080/00273170802034810

Beisly, A., Kwon, K.-A., & Jeon, S. (2020). Executive function and learning behaviors: Associations with academic skills among preschoolers. *Early Child Development and Care*, *190*(15), 2469–2483. https://doi.org/10.1080/03004430.2019.1585347

Benjamini, Y., & Hochberg, Y. (1995). controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function: Development of executive functions. *Child Development*, *81*(6), 1641–1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences*, *21*(4), 327–336. https://doi.org/10.1016/j.lindif.2011.01.007

Bierman, K. L., Nix, R. L., Greenberg, M. T., Blair, C., & Domitrovich, C. E. (2008). Executive functions and school readiness intervention: Impact, moderation, and mediation in the Head Start REDI program. *Development and Psychopathology*, *20*(3), 821–843. https://doi.org/10.1017/S0954579408000394

Blair, C., & McKinnon, R. D. (2016). Moderating effects of executive functions and the teacher–child relationship on the development of mathematics ability in kindergarten. *Learning and Instruction*, *41*, 85–93. https://doi.org/10.1016/j.learninstruc.2015.10.001

Box, G. E. P. (1954). Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, II. Effects of Inequality of Variance and of Correlation Between Errors in the Two-Way Classification. *The Annals of Mathematical Statistics*, *25*(3), 484–498. http://www.jstor.org/stable/2236831

Bryant, D. P., Bryant, B. R., Gersten, R., Scammacca, N., & Chavez, M. M. (2008). Mathematics intervention for first- and second-grade students with mathematics difficulties: The effects of tier 2 intervention delivered as booster lessons. *Remedial and Special Education*, *29*(1), 20–32. https://doi.org/10.1177/0741932507309712

Bryant, D. P., Bryant, B. R., Roberts, G., Vaughn, S., Pfannenstiel, K. H., Porterfield, J., & Gersten, R. (2011). Early numeracy intervention program for first-grade students with mathematics difficulties. *Exceptional Children*, *78*(1), 7–23. https://doi.org/10.1177/001440291107800101

Brydges, C. R., Reid, C. L., Fox, A. M., & Anderson, M. (2012). A unitary executive function predicts intelligence in children. *Intelligence*, *40*(5), 458–469. https://doi.org/10.1016/j.intell.2012.05.006

Butler, F. M., Miller, S. P., Crehan, K., Babbitt, B., & Pierce, T. (2003). Fraction instruction for students with mathematics disabilities: Comparing two teaching sequences. *Learning Disabilities Research and Practice*, *18*(2), 99–111. https://doi.org/10.1111/1540-5826.00066

Clark, C. A. C., Pritchard, V. E., & Woodward, L. J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Developmental Psychology*, *46*(5), 1176–1191. https://doi.org/10.1037/a0019672

Clarke, B., Doabler, C., Fien, H., Baker, S. K., & Smolkowski, K. (2011, September). *A Randomized control trial of a tier 2 kindergarten mathematics intervention.* (IES Grant No. R324A120304). Eugene, OR: University of Oregon.

Clarke, B., Doabler, C., Fien, H., & Smolkowski, K. (2015, August). *A randomized control trial of a Tier 2 first-grade mathematics intervention.* (IES Grant No. R324A160046). Eugene, OR: University of Oregon.

Clarke, B., Doabler, C. T., Smolkowski, K., Baker, S. K., Fien, H., & Strand Cary, M. (2016). Examining the efficacy of a tier 2 kindergarten mathematics intervention. *Journal of Learning Disabilities*, *49*(2), 152–165. https://doi.org/10.1177/0022219414538514

Clarke, B., Doabler, C. T., Smolkowski, K., Turtura, J., Kosty, D., Kurtz-Nelson, E., Fien, H., & Baker, S. K. (2019). Exploring the relationship between initial mathematics skill and a kindergarten mathematics intervention. *Exceptional Children*, *85*(2), 129–146.

Clarke, B., Doabler, C. T., Strand Cary, M., Kosty, D., Baker, S., Fien, H., & Smolkowski, K. (2014). Preliminary evaluation of a tier 2 mathematics intervention for first-grade students: Using a theory of change to guide formative evaluation activities. *School Psychology Review*, *43*(2), 160–178. https://doi.org/10.1080/02796015.2014.12087442

Clarke, B., Doabler, C. T., Sutherland, M., Kosty, D., Turtura, J., & Smolkowski, K. (in press). Examining the impact of a first grade whole number intervention by group size. *Journal of Research on Educational Effectiveness.*

Clarke, B., Doabler, C. T., Turtura, J., Smolkowski, K., Kosty, D. B., Sutherland, M., Kurtz-Nelson, E., Fien, H., & Baker, S. K. (2020). Examining the efficacy of a kindergarten mathematics intervention by group size and initial skill: Implications for practice and policy. *The Elementary School Journal*, *121*(1), 125–153. https://doi.org/10.1086/710041

Clarke, B., Gersten, R. M., Dimino, J., & Rolfhus, E. (2011). *Assessing student proficiency of number sense (ASPENS)* [Measurement instrument]. Cambium Learning Group, Sopris Learning.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155.

Common Core State Standards Initiative. (2010). Common core standards for mathematics. Retrieved 12/15/10, from http://www.corestandards.org/the-standards/mathematics

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education*, *3*(2), 63–68. https://doi.org/10.1016/j.tine.2013.12.001

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Doabler, C. T., Clarke, B., Kosty, D., Kurtz-Nelson, E., Fien, H., Smolkowski, K., & Baker, S. K. (2019). Examining the impact of group size on the treatment intensity of a tier 2 mathematics intervention within a systematic framework of replication. *Journal of Learning Disabilities*, *52*(2), 168–180. https://doi.org/10.1177/0022219418789376

Doabler, C. T., Clarke, B., Kosty, D., Turtura, J. E., Sutherland, M., Maddox, S. A., & Smolkowski, K. (2021). Using direct observation to document "practice-based evidence" of evidence-based mathematics instruction. *Journal of Learning Disabilities*, *54*(1), 20–35. https://doi.org/10.1177/0022219420911375

Doabler, C. T., & Fien, H. (2013). Explicit mathematics instruction: What teachers can do for teaching students with mathematics difficulties. *Intervention in School and Clinic*, *48*(5), 276–285. https://doi.org/10.1177/1053451212473151

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., & Duckworth, K. (2007). School readiness and later achievement. *Developmental Psychology*, *46*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

Dyson, N. I., Jordan, N. C., & Glutting, J. (2011). A number sense intervention for low-income kindergartners at risk for mathematics difficulties. *Journal of Learning Disabilities*, *46*(2), 166–181. https://doi.org/10.1177/0022219411410233

Evers, W. F., Walk, L. M., Quante, S., & Hille, K. (2016). Relations between measures of executive functions and self-regulation in preschoolers. *Diskurs Kindheits- Und Jugendforschung*, *11*(4), 435–455. https://doi.org/10.3224/diskurs.v11i4.25603

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

Frye, D., Baroody, A. J., Burchinal, M., Carver, S. M., Jordan, N. C., & McDowell, J. (2013). *Teaching math to young children: A practice guide* (No. NCEE 2014-4005). National Center for Education and Regional Assistance (NCEE), Institute of Education Sciences, US Department of Education. https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/early_math_pg_111313.pdf

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97*, 493–513. https://doi.org/10.1037/0022-0663.97.3.493

Fuchs, D., & Fuchs, L. S. (2019). On the importance of moderator analysis in intervention research: An introduction to the special issue. *Exceptional Children, 85*(2), 126–128. https://doi.org/10 .1177/0014402918811924

Fuchs, L. S., Fuchs, D., & Gilbert, J. K. (2019). Does the severity of students' pre-intervention math deficits affect responsiveness to generally effective first-grade intervention? *Exceptional Children*, *85*(2), 147–162. https://doi.org/10.1177/0014402918782628

Fuchs, L. S., Malone, A. S., Schumacher, R. F., Namkung, J., Hamlett, C. L., Jordan, N. C., Siegler, R. S., Gersten, R., & Changas, P. (2016). Supported self-explaining during fraction intervention. *Journal of Educational Psychology*, *108*, 493–508. https://doi.org/10.1037/edu0000073

Fuchs, L.S., Newman-Gonchar, R., Schumacher, R., Dougherty, B., Bucka, N., Karp, K.S., Woodward, J., Clarke, B., Jordan, N. C., Gersten, R., Jayanthi, M., Keating, B., and Morgan, S. (2021). Assisting Students Struggling with Mathematics: Intervention in the Elementary Grades (WWC 2021006). Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Retrieved from http://whatworks.ed.gov/.

Fuchs, L. S., Schumacher, R. F., Sterba, S. K., Long, J., Namkung, J., Malone, A., Hamlett, C. L., Jordan, N. C., Gersten, R., Siegler, R. S., & Changas, P. (2014). Does working memory moderate the effects of fraction intervention? An aptitude–treatment interaction. *Journal of Educational Psychology*, *106*(2), 499–514. https://doi.org/10.1037/a0034341

Fuchs, L. S., Sterba, S. K., Fuchs, D., & Malone, A. S. (2016). Does evidence-based fractions intervention address the needs of very low-performing students? *Journal of Research on Educational Effectiveness*, *9*, 662–677. https://doi.org/10.1080/19345747.2015.1123336

Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-intervention: A decade later. *Journal of Learning Disabilities*, *45*(3), 195–203. https://doi.org/10.1177/0022219412442150

Fuhs, M. W., Nesbitt, K. T., Farran, D. C., & Dong, N. (2014). Longitudinal associations between executive functioning and academic skills across content areas. *Developmental Psychology*, *50*(6), 1698–1709. https://doi.org/10.1037/a0036633

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research*, *79*(3), 1202–1242. https://doi.org/10.3102/0034654309334431

Gersten, R., Rolfhus, E., Clarke, B., Decker, L. E., Wilkins, C., & Dimino, J. (2015). intervention for first graders with limited number knowledge: Large-scale replication of a randomized controlled trial. *American Educational Research Journal*, *52*(3), 516–546. https://doi.org/10.3102/0002831214565787

Ginsburg, H. P., & Baroody, A. J. (2003). *Test of early mathematics ability- Third edition (TEMA-3)*. ProEd.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*(1), 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Hassinger-Das, B., Jordan, N. C., Glutting, J., Irwin, C., & Dyson, N. (2014). Domain-general mediators of the relation between kindergarten number sense and first-grade mathematics achievement. *Journal of Experimental Child Psychology*, *118*, 78–92. https://doi.org/10.1016/j.jecp.2013.09.008

Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, *19*(4), 451–464. https://doi.org/10.1016/j.labeco.2012.05.014

IBM Corp. (2019). IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp. [Computer Software].

Individuals with Disabilities Education Act of 2004, Public Law 108-446, 34 C.F.R. § 300.8 (c)(10) (2004).

Jacob, R., & Parkinson, J. (2015). The potential for school-based interventions that target executive function to improve academic achievement: A review. *Review of Educational Research*, *85*(4), 512–552. https://doi.org/10.3102/0034654314561338

Jimerson, S., Egeland, B., & Teo, A. (1999). A longitudinal study of achievement trajectories. *Journal of Educational Psychology*, *91*(1), 116–126. https://doi.org/10.1037/0022-0663.91.1.116

Jordan, N. C., Glutting, J., & Ramineni, C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learning and Individual Differences*, *20*(2), 82–88. https://doi.org/10.1016/j.lindif.2009.07.004

Jordan, N. C., Kaplan, D., Nabors Olah, L., & Locuniak, M. N. (2006). number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development*, *77*(1), 153–175. https://doi.org/10.1111/j.1467-8624.2006.00862.x

Jung, P.-G., McMaster, K. L., Kunkel, A. K., Shin, J., & Stecker, P. M. (2018). Effects of data-based individualization for students with intensive learning needs: A meta-analysis: Effects of data-based individualization. *Learning Disabilities Research & Practice*, *33*(3), 144–155. https://doi.org/10.1111/ldrp.12172

Keller, L., & Libertus, M. (2015). Inhibitory control may not explain the link between approximation and math abilities in kindergarteners from middle class families. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00685

Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development*, *67*(2), 490–507. https://doi.org/10.1111/j.1467-8624.1996.tb01747.x

Kroesbergen, E. H., van 't Noordende, J. E., & Kolkman, M. E. (2014). Training working memory in kindergarten children: Effects on working memory and early numeracy. *Child Neuropsychology*, *20*(1), 23–37. https://doi.org/10.1080/09297049.2012.736483

Lan, X., Legare, C. H., Ponitz, C. C., Li, S., & Morrison, F. J. (2011). Investigating the links between the subcomponents of executive function and academic achievement: A cross-cultural analysis of Chinese and American preschoolers. *Journal of Experimental Child Psychology*, *108*(3), 677–692. https://doi.org/10.1016/j.jecp.2010.11.001

LeFevre, J.-A., Berrigan, L., Vendetti, C., Kamawar, D., Bisanz, J., Skwarchuk, S.-L., & Smith-Chant, B. L. (2013). The role of executive attention in the acquisition of mathematical skills for children in Grades 2 through 4. *Journal of Experimental Child Psychology*, *114*(2), 243–261. https://doi.org/10.1016/j.jecp.2012.10.005

Lerner, M. D., & Lonigan, C. J. (2014). Executive function among preschool children: Unitary versus distinct abilities. *Journal of Psychopathology and Behavioral Assessment, 36*(4), 626–639. https://doi.org/10.1007/s10862-014-9424-3

Liew, J. (2012). Effortful control, executive functions, and education: bringing self-regulatory and social-emotional competencies to the table: Self-regulation and education. *Child Development Perspectives*, *6*(2), 105–111. https://doi.org/10.1111/j.1750-8606.2011.00196.x

Lin, X., & Powell, S. R. (2021). The roles of initial mathematics, reading, and cognitive skills in subsequent mathematics performance: A meta-analytic structural equation modeling approach. *Review of Educational Research*, 3465432110545. https://doi.org/10.3102/00346543211054576

Lonigan, C. J., Allan, D. M., Goodrich, J. M., Farrington, A. L., & Phillips, B. M. (2017). Inhibitory control of spanish-speaking language-minority preschool children: Measurement and association with language, literacy, and math skills. *Journal of Learning Disabilities*, *50*(4), 373–385. https://doi.org/10.1177/0022219415618498

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly*, *21*(4), 471–490. https://doi.org/10.1016/j.ecresq.2006.09.003

McClelland, M. M., Acock, A. C., Piccinin, A., Rhea, S. A., & Stallings, M. C. (2013). Relations between preschool attention span-persistence and age 25 educational outcomes. *Early Childhood Research Quarterly*, *28*(2), 314–324. https://doi.org/10.1016/j.ecresq.2012.07.008

McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology*, *43*(4), 947–959. https://doi.org/10.1037/0012-1649.43.4.947

McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., & Pratt, M. E. (2014). Predictors of early growth in academic achievement: The head-toes-knees-shoulders task. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00599

McClelland, M. M., Gonzales, C. R., Cameron, C. E., Geldhof, G. J., Bowles, R. P., Nancarrow, A. F., Merculief, A., & Tracy, A. (2021). The head-toes-knees-shoulders revised: Links to academic outcomes and measures of EF in young children. *Frontiers in Psychology*, *12*, 721846. https://doi.org/10.3389/fpsyg.2021.721846

McKinnon, R. D., & Blair, C. (2019). Bidirectional relations among executive function, teacher–child relationships, and early reading and math achievement: A cross-lagged panel analysis. *Early Childhood Research Quarterly*, *46*, 152–165. https://doi.org/10.1016/j.ecresq.2018.03.011

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*(2), 270–291. https://doi.org/10.1037/a0028228

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of "far transfer": Evidence from a meta-analytic review. *Perspectives on Psychological Science*, *11*(4), 512–534. https://doi.org/10.1177/1745691616635612

Mellard, D., McKnight, M., & Jordan, J. (2010). RTI tier structures and instructional intensity. *Learning Disabilities Research & Practice*, *25*(4), 217–225. https://doi.org/10.1111/j.1540-5826.2010.00319.x

Miller, B., Vaughn, S., & Freund, L. S. (2014). Learning disabilities research studies: Findings from NICHD-funded projects. *Journal of Research on Educational Effectiveness*, *7*(3), 225–231. https://doi.org/10.1080/19345747.2014.927251

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Morgan, P. L., Farkas, G., Wang, Y., Hillemeier, M. M., Oh, Y., & Maczuga, S. (2019). Executive function deficits in kindergarten predict repeated academic difficulties across elementary school. *Early Childhood Research Quarterly*, *46*, 20–32. https://doi.org/10.1016/j.ecresq.2018.06.009

Morgan, P. L., Farkas, G., & Wu, Q. (2011). Kindergarten children's growth trajectories in reading and mathematics: Who falls increasingly behind? *Journal of Learning Disabilities*, *44*(5), 472–488. https://doi.org/10.1177/0022219411414010

Murray, D. M. (1998). *Design and analysis of group-randomized trials.* Oxford University Press.

National Center for Education Statistics. (2019). The nation's report card. Available from https://nces.ed.gov/nationsreportcard/mathematics

Nelson, G., & Powell, S. R. (2018). A systematic review of longitudinal studies of mathematics difficulty. *Journal of Learning Disabilities*, *51*(6), 523–539. https://doi.org/10.1177/0022219417714773

Nemmi, F., Helander, E., Helenius, O., Almeida, R., Hassler, M., Räsänen, P., & Klingberg, T. (2016). Behavior and neuroimaging at baseline predict individual response to combined mathematical and working memory training in children. *Developmental Cognitive Neuroscience*, *20*, 43–51. https://doi.org/10.1016/j.dcn.2016.06.004

Nguyen, T., & Duncan, G. J. (2019). Kindergarten components of executive function and third grade achievement: A national study. *Early Childhood Research Quarterly*, *46*, 49–61. https://doi.org/10.1016/j.ecresq.2018.05.006

Nguyen, T., Duncan, R. J., & Bailey, D. H. (2019). Theoretical and methodological implications of associations between executive function and mathematics in early childhood. *Contemporary Educational Psychology*, *58*, 276–287. https://doi.org/10.1016/j.cedpsych.2019.04.002

Nigg, J. T. (2017). Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry*, *58*(4), 361–383. https://doi.org/10.1111/jcpp.12675

Ntourou, K., Anderson, J. D., & Wagovich, S. A. (2018). Executive function and childhood stuttering: Parent ratings and evidence from a behavioral task. *Journal of Fluency Disorders*, *56*, 18–32. https://doi.org/10.1016/j.jfludis.2017.12.001

Peng, P., & Miller, A. C. (2016). Does attention training work? A selective meta-analysis to explore the effects of attention training and moderators. *Learning and Individual Differences*, *45*, 77–87. https://doi.org/10.1016/j.lindif.2015.11.012

Ponitz, C. C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental Psychology*, *45*(3), 605–619. https://doi.org/10.1037/a0015365

Powell, S. R., Cirino, P. T., & Malone, A. S. (2017). Child-level predictors of responsiveness to evidence-based mathematics intervention. *Exceptional Children*, *83*(4), 359–377. https://doi.org/10.1177/0014402917690728

Prager, E. O., Sera, M. D., & Carlson, S. M. (2016). Executive function and magnitude skills in preschool children. *Journal of Experimental Child Psychology*, *147*, 126–139. https://doi.org/10.1016/j.jecp.2016.01.002

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, *31*(4), 437–448. https://doi.org/10.3102/10769986031004437

Purpura, D. J., Schmitt, S. A., & Ganley, C. M. (2017). Foundations of mathematics and literacy: The role of executive functioning components. *Journal of Experimental Child Psychology*, *153*, 15–34. https://doi.org/10.1016/j.jecp.2016.08.010

Rabiner, D. L., Godwin, J., & Dodge, K. A. (2016). predicting academic achievement and attainment: The contribution of early academic skills, attention difficulties, and social competence. *School Psychology Review*, *45*(2), 250–267. https://doi.org/10.17105/SPR45-2.250-267

Raghubar, K., Cirino, P., Barnes, M., Ewing-Cobbs, L., Fletcher, J., & Fuchs, L. (2009). Errors in multi-digit arithmetic and behavioral inattention in children with math difficulties. *Journal of Learning Disabilities*, *42*(4), 356–371. https://doi.org/10.1177/0022219409335211

Ribner, A. D. (2020). Executive function facilitates learning from math instruction in kindergarten: Evidence from the ECLS-K. *Learning and Instruction*, *65*, 101251. https://doi.org/10.1016/j.learninstruc.2019.101251

Ribner, A. D., Willoughby, M. T., & Blair, C. B. (2017). Executive function buffers the association between early math and later academic skills. *Frontiers in Psychology*, *8*, 869. https://doi.org/10.3389/fpsyg.2017.00869

Ritchie, S. J., & Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological Science*, *24*(7), 1301–1308. https://doi.org/10.1177/0956797612466268

SAS Institute. (2016). *Base SAS® 9.4 procedures guide: Statistical procedures (5th ed.).* SAS Institute, Inc.

Schmitt, S. A., Geldhof, G. J., Purpura, D. J., Duncan, R., & McClelland, M. M. (2017). Examining the relations between executive function, math, and literacy during the transition to kindergarten: A multi-analytic approach. *Journal of Educational Psychology*, *109*(8), 1120–1140. https://doi.org/10.1037/edu0000193

Schmitt, S. A., Pratt, M. E., & McClelland, M. M. (2014). Examining the validity of behavioral self-regulation tools in predicting preschoolers' academic achievement. *Early Education and Development*, *25*(5), 641–660. https://doi.org/10.1080/10409289.2014.850397

Schulte, A. C., & Stevens, J. J. (2015). Once, sometimes, or always in special education: Mathematics growth and achievement gaps. *Exceptional Children*, *81*(3), 370–387. https://doi.org/10.1177/0014402914563695

Schumacher, R. F., Zumeta Edmonds, R., & Arden, S. V. (2017). Examining implementation of intensive intervention in mathematics: Learning disabilities research. *Learning Disabilities Research & Practice*, *32*(3), 189–199. https://doi.org/10.1111/ldrp.12141

Shanley, L., Clarke, B., Smolkowski, K., Doabler, C. T., Kurtz-Nelson, E. C., & Fien, H. (2021). Examining the role of domain-general skills in mathematics learning and intervention response in kindergarten. *Learning Disabilities Research & Practice*, *36*(4), 330–352. https://doi.org/10.1111/ldrp.12267

Smith, J. L. M., Sáez, L., & Doabler, C. T. (2016). Using explicit and systematic instruction to support working memory. *TEACHING Exceptional Children*, *48*(6), 275–281. https://doi.org/10.1177/0040059916650633

Sung, J., & Wickrama, K. A. S. (2018). Longitudinal relationship between early academic achievement and executive function: Mediating role of approaches to learning. *Contemporary Educational Psychology*, *54*, 171–183. https://doi.org/10.1016/j.cedpsych.2018.06.010

Titz, C., & Karbach, J. (2014). Working memory and executive functions: Effects of training on academic achievement. *Psychological Research*, *78*(6), 852–868. https://doi.org/10.1007/s00426-013-0537-1

Toll, S. W. M., & Van Luit, J. E. H. (2012). Early numeracy intervention for low-performing kindergartners. *Journal of Early Intervention*, *34*(4), 243–264. https://doi.org/10.1177/1053815113477205

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner Review: Do performance-based measures and ratings of executive function assess the same construct?: Performance-based and rating measures of EF. *Journal of Child Psychology and Psychiatry*, *54*(2), 131–143. https://doi.org/10.1111/jcpp.12001

U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Mathematics Assessment.

van Belle. (2008). *Statistical rules of thumb* (2nd ed.). Wiley.

Vaughn, S., Linan-Thompson, S., Kouzekanani, K., Pedrotty Bryant, D., Dickson, S., & Blozis, S. A. (2003). Reading instruction grouping for students with reading difficulties. *Remedial and Special Education*, *24*(5), 301–315. https://doi.org/10.1177/07419325030240050501

Vaughn, S., & Swanson, E. A. (2015). Special education research advances knowledge in education. *Exceptional Children*, *82*(1), 11–24. https://doi.org/10.1177/0014402915598781

Viterbori, P., Usai, M. C., Traverso, L., & De Franchis, V. (2015). How preschool executive functioning predicts several aspects of math achievement in Grades 1 and 3: A longitudinal study. *Journal of Experimental Child Psychology*, *140*, 38–55. https://doi.org/10.1016/j.jecp.2015.06.014

Wang, Z. & Thompson, B. (2007). Is the pearson r^2 biased, and if so, what is the best correction formula? *Journal of Experimental Education*, *75*(2), 109-125.

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review, 36*, 541–561.

Wass, S. V., Scerif, G., & Johnson, M. H. (2012). Training attentional control and working memory – Is younger, better? *Developmental Review*, *32*(4), 360–387. https://doi.org/10.1016/j.dr.2012.07.001

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology*, *102*(1), 43–53. https://doi.org/10.1037/a0016738

Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A. C., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology*, *108*(3), 436–452. https://doi.org/10.1016/j.jecp.2010.08.008

Willoughby, M. T., Blair, C. B., Wirth, R. J., Greenberg, M., & The Family Life Project Investigators. (2012). The measurement of executive function at age 5: Psychometric properties and relationship to academic achievement. *Psychological Assessment*, *24*(1), 226–239. https://doi.org/10.1037/a0025361

Zhou, Q., Chen, S. H., & Main, A. (2012). Commonalities and differences in the research on children’s effortful control and executive function: a call for an integrated model of self-regulation: Effortful control and executive function. *Child Development Perspectives*, *6*(2), 112–121. https://doi.org/10.1111/j.1750-8606.2011.00176.x